Journal of Medical Internet Research

Journal Impact Factor (JIF) (2022): 7.4
Volume 22 (2020), Issue 5 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

Original Papers

Viewpoints

Tutorial

Short Paper

Corrigenda and Addenda

Proposal

Review

Patients' Experiences of Telehealth in Palliative Home Care: Scoping Review

Simen A Steindal[1], PhD; Andréa Aparecida Goncalves Nes[1], PhD; Tove E Godskesen[2,3], PhD; Alfhild Dihle[4], PhD; Susanne Lind[2], PhD; Anette Winger[4], PhD; Anna Klarare[2,5], PhD

[1]Lovisenberg Diaconal University College, Oslo, Norway

[2]Palliative Research Centre, Department of Health Care Sciences, Ersta Sköndal Bräcke University College, Stockholm, Sweden

[3]Centre for Research Ethics & Bioethics, Uppsala University, Uppsala, Sweden

[4]Oslo Metropolitan University, Oslo, Norway

[5]Clinical Psychology in Healthcare, Department for Women's and Children's Health, Uppsala University, Uppsala, Sweden

Corresponding Author:
Simen A Steindal, PhD
Lovisenberg Diaconal University College
Lovisenberggt 15B
Oslo
Norway
Phone: 47 92660422
Email: simen.alexander.steindal@ldh.no

Abstract

Background: Telehealth is increasingly being used in home care and could be one measure to support the needs of home-based patients receiving palliative care. However, no previous scoping review has mapped existing studies on the use of telehealth for patients in palliative home care.

Objective: The aim of this study was to map and assess published studies on the use of telehealth for patients in palliative home care.

Methods: A scoping review was conducted using the methodological framework of Arksey and O'Malley. Reporting was guided by Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews. A systematic and comprehensive search of Medical Literature Analysis and Retrieval System Online, EMBASE, PsycINFO, and Cumulative Index to Nursing and Allied Health was performed for studies published between January 2000 and October 2018. Two authors independently assessed eligibility and extracted data.

Results: The review included 22 papers from 19 studies. Four thematic groupings were identified among the included papers: easy and effortless use of telehealth regardless of the current health condition, visual features that enhance communication and care via telehealth, symptom management and self-management promotion by telehealth, and perceptions of improved palliative care at home.

Conclusions: The use of telehealth in palliative home care seems to be feasible, improving access to health care professionals at home and enhancing feelings of security and safety. The visual features of telehealth seem to allow a genuine relationship with health care professionals. However, there are contradicting results on whether the use of telehealth improves burdensome symptoms and quality of life. Future research should investigate the experiences of using telehealth among patients with life-limiting illness other than cancer and patients aged 85 years or older. More research is needed to increase the body of knowledge regarding the effectiveness of telehealth on symptoms and quality of life.

(J Med Internet Res 2020;22(5):e16218) doi:10.2196/16218

KEYWORDS

health care technology; review; palliative care; telemedicine

Introduction

The preferred place of care for most patients in need of palliative care is their own home, and many of them are able to spend time at home and receive the needed care [1,2]. Patients want, as usual, to feel meaning and to maintain governance of their lives; thus, they prefer independent access to health care professionals when needed. Moreover, patients in need of palliative care express preferences for continuity of care and for health care professionals to provide coordinated care [2,3], which may be challenging in home care. Patients who receive care at home report unmet palliative care needs, such as the lack of regular communication with nurses and physicians and between primary and secondary health care professionals [4]. Furthermore, patients may feel uncertain about the urgency of their problems, whom they should contact in times of need, what the response may be, and the legitimacy of their needs, in addition to perceptions of poor continuity in home follow-up and coordination of services [4-6]. Physicians, nurses, physiotherapists, and social workers are often involved in the care and follow-up of patients in need of palliative care at home [7].

Telehealth is increasingly being used in home care [8] and could be one measure to meet the reported challenges and support the needs of home-based patients receiving palliative care. Telehealth is defined as "the provision of healthcare remotely by means of a variety of telecommunication tools" [8]. Telehealth can be delivered in an interactive mode, which invites an exchange of information or messages between patients and health care professionals, or in a passive mode, which is a form of communication that does not require an immediate response by the recipient [9].

Telehealth may be useful for conditions that require close monitoring, clinical assessment, and early intervention to prevent adverse events, such as unwanted emergency hospitalization [10]. The potential benefits of telehealth could include increased quality of life by improving independence and self-management with increased choice, improved access to community palliative care services for those wishing to die at home, and reduction in unnecessary hospital admissions [11]. Moreover, telehealth could be used to reduce patient travel burdens and provide access to services after regular clinic hours [12]. The use of telehealth appears to be promising as a help to meet patients' expectations and needs related to maintenance of their care at home. However, challenges also exist. Head et al [13] claimed that palliative care has been characterized as high touch rather than high tech, which could limit the interest of health care professionals in applying technological advancements when developing and honing interventions [14].

Several systematic reviews have examined home-based telehealth in palliative care settings. One systematic review examined the evidence for home-based telehealth in pediatric care by including studies identified in two databases, focusing on children, adults, and health care professionals [14]. Other systematic reviews have examined patient outcomes [13] and caregiver outcomes [14,15] on telehealth interventions. Head et al [13] found heterogeneity regarding patient population and technology use. The outcome measures showed that all the included studies, except one, reported improvement in quality of life or symptom management. Another systematic literature review assessed the effectiveness of electronic health interventions for patients in palliative care or stakeholders such as health care professionals or caregivers [16]. Some of the included studies indicated positive results regarding quality of care, communication, and cost development. However, no randomized controlled trials (RCTs) were found.

Other literature reviews have been limited to geographical areas and have explored the use of telehealth in palliative care in the United Kingdom [17,18] and the development and use of mobile devices in palliative care services in sub-Saharan Africa [19]. In addition, the review by Johnston [17] was limited to older people. Kidd et al [18] found that telehealth was used by patients, relatives, and health care professionals in several contexts related to palliative care: oncology settings, specialist palliative care, primary care, and nursing homes. Telehealth applications including videoconferencing; consultations; symptom assessments; and advice for patients, relatives, and health care professionals were deemed usable and acceptable for patients and health care professionals.

Telehealth is increasingly being used in patients' homes [8], and new technologies are being developed and implemented rapidly. Furthermore, as previously mentioned, some of the previous literature reviews are older and limited by geographical areas and to older patients. Consequently, there is a need to conduct an updated and broader literature review to develop an overview of the body of knowledge within this field and to identify gaps in knowledge for evidence-based practice. To our knowledge, no scoping review has mapped existing studies on the use of telehealth for patients in palliative home care. Consequently, this scoping review aimed to map and assess published studies on the use of telehealth for patients in palliative home care. The specific research question was as follows: What is known from the existing research literature about patients' experiences of the use of telehealth in palliative home care?

Methods

Design

This scoping review used the framework of Arksey and O'Malley [20], which comprises five stages: identifying the research question; identifying relevant studies; study selection; charting the data; and collating, summarizing, and reporting the results. The reporting of the scoping review was guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) checklist [21]. The protocol for this scoping review has not been registered or published.

Identifying Relevant Studies

A systematic broad search was performed in October 2018 using the databases Medical Literature Analysis and Retrieval System Online (MEDLINE), PsycINFO, EMBASE, and Cumulative Index to Nursing and Allied Health (CINAHL) for studies published between January 1, 2000, and October 16, 2018. The

search strategy was built in MEDLINE by 4 of the authors (SS, AAGN, AW, and AK) and a librarian using Medical Subjects Headings and text words. The search was adopted for each subsequent database. The search strategy is described in Multimedia Appendix 1. In addition, a hand search was performed to screen the reference lists of the included papers.

Study Selection

On the basis of the inclusion and exclusion criteria (see Table 1), pairs of authors independently screened titles, abstracts, and full-text papers for inclusion in the study. When there was disagreement, an independent assessment of whether or not a publication met the inclusion criteria was conducted by a third author.

Table 1. Inclusion and exclusion criteria.

Criterion	Inclusion	Exclusion
Types of studies	Qualitative, quantitative, and mixed method studies on the phenomenon published in peer-reviewed journals	Letters, comments, conference abstracts, editorials, doctoral thesis, or any type of review
Period	January 1, 2000, until October 16, 2018	Before January 1, 2000, and after October 16, 2018
Language	English, Portuguese, Spanish, or Scandinavian	All other languages
Type of participants	Patients in a palliative care trajectory regardless of diagnosis, aged 18 years or older, and living at home	Patients who are not in a palliative care trajectory; patients aged 17 years or younger; and patients who use telehealth in a hospital, nursing home, or hospice setting
Phenomenon of interest	Patients' experiences of using telehealth at home with follow-up from health care professionals	Patients' experiences of using telehealth at home without follow-up from health care professionals or experiences of using telehealth in a hospital, nursing home, or hospice setting
Type of outcomes	Patients reported subjective and objective outcomes	Proxy-reported (next of kin or health care professional) outcomes

Charting the Data

Pairs of authors extracted data from the included publications, using a standardized data-charting form and maintaining the wording and terminology from the papers. The form included the following information: authors, year of publication, country of origin, aim, population and sample size, telehealth application, delivered mode, design and method, and results (see Multimedia Appendix 2). Any disagreements were resolved by a third author.

Collating, Summarizing, and Reporting the Results

An inductive approach was used to thematically organize and summarize the results from the included papers to answer the research question [20]. The extracted results from each paper were read several times to identify frequent patterns, similarities, and differences in patients' experiences of using telehealth, regardless of the type of technology. The identified emerging patterns were organized in four thematic groupings. The first, second, and last author discussed the results and agreed upon the final groupings of the results. A frequency table illustrating which articles were included in which grouping was made (see Table 2).

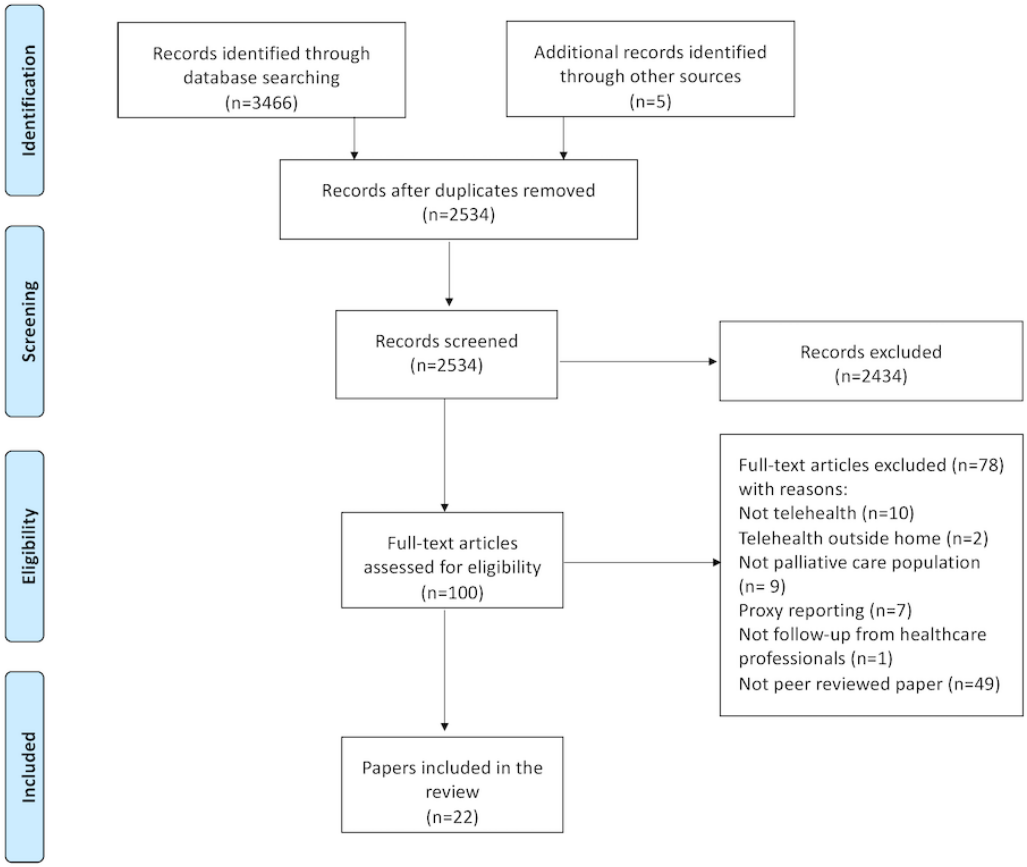
Table 2. Articles included in thematic groupings.

Theme	Study	Number of articles
Easy and effortless use regardless of the current health condition	Aoki et al [22], Hochstenbach et al [23], McCall et al [24], Pinto et al [25], Whitten et al [26], Besse et al [27], Lind and Karlsson [28], Passik et al [29], Tieman et al [30], Reinke et al [31], Lind [32], Lind et al [33], Stern et al [34], Hennemann-Krause et al [35], van Gurp et al [36]	15
Visual features enhance communication and care via telehealth	Hebert et al [37], Miyazaki et al [38], Whitten et al [39], Whitten et al [26], Passik et al [29], Wilkinson et al [40], Reinke et al [31], Stern et al [34], Hennemann-Krause et al [35], van Gurp [36]	10
Symptom management and self-management promotion by telehealth	Bonsignore et al [41], Hebert et al [37], Hochstenbach et al [23], McCall et al [24], Miyazaki et al [38], Pinto et al [25], Whitten et al 30], Besse et al [27], Hoek et al [42], Lind and Karlsson [28], Wilkinson et al [40], Lind [32], Lind et al [33], Hennemann-Krause et al [35]	14
Perceptions of improved palliative care at home	Aoki et al [22], Bonsignore et al [41], Hochstenbach et al [23], McCall et al [24], Pinto et al [25], Whitten et al [39], Whitten et al [26], Lind and Karlsson [28], Wilkinson et al [40], Lind [32], Lind et al [33], Stern et al [34], van Gurp et al [43], van Gurp et al [36]	14

Results

The database and hand searches yielded 3471 publications. After 937 duplicates were removed, titles and abstracts for 2532 publications were screened. On the basis of the inclusion and exclusion criteria, full text of 100 publications were read; 78 publications were excluded and 22 publications from 19 studies were included in the review (see Figure 1).

Figure 1. Summary of the selection of studies.



Description of Included Studies

A total of 22 papers from 19 studies were included. The included studies were conducted in Australia (n=1), Brazil (n=1), Canada (n=2), Japan (n=1), the Netherlands (n=4), Portugal (n=1), Sweden (n=1), United Kingdom (n=3), and United States (n=5). The sample size of the included studies ranged from 2 to 187 participants, and in 6 papers, the samples consisted of 11 participants or fewer. Papers included patients with cancer (n=13); chronic obstructive pulmonary disease (COPD; n=1); cystic fibrosis (n=1); and a mix of different life-limiting illnesses such as cancer, COPD, multiple sclerosis, and amyotrophic lateral sclerosis (n=4). Three papers did not report diagnoses. Six papers included patients aged 85 years and above (oldest-old patients). Nine papers used mixed method or multimethod design, combining qualitative and quantitative methods [22-26]; 7 papers used quantitative design [27-31]; 3 papers used case study design [32-34]; 1 paper used case series design [35]; and 2 papers used qualitative design [36,43]. Three papers included an RCT [37,40,42]. The characteristics of the included studies are shown in Multimedia Appendix 2.

In 15 papers, telehealth was delivered using an interactive mode, whereas the passive mode was used in 7 papers (see Multimedia Appendix 2).

Video-based technology (n=14) was the most frequently used telehealth application in palliative home care. In 10 papers, teleconsultation with patients, relatives, and health care professionals was used to discuss patients' needs, concerns, symptoms, and other problems and to give patients advice [22,26,30,35,36,39-42]. When possible, patients' general practitioners participated from patients' homes [36,42,43]. In four papers, videotelephone was used for individual contact between patients and health care professionals for support, symptom management [34,37,38], and dignity psychotherapy [29]. Webinar technology without video was used to enhance patients' knowledge and skills about end-of-life issues and conversations [31].

Mobile devices such as mobile phones [27]; PDAs [24]; apps on smartphones, tablets, or PCs [23,25,41]; and digital pens and diaries [28,32] were used for pain education and the monitoring and management of pain and other symptoms.

Two studies described a theoretical framework for telehealth intervention: self-monitoring [23] and dignity psychotherapy [29].

To answer the research question regarding what is known about patients' experiences of the use of telehealth in palliative home care, the results of this scoping review are presented in four thematic groupings: easy and effortless use of telehealth regardless of the current health condition, visual features that enhance communication and care via telehealth, symptom management and self-management promotion by telehealth, and perceptions of improved palliative care at home (see Table 2).

Easy and Effortless Use of Telehealth Regardless of the Current Health Condition

Fifteen papers reported patients' experiences of using the telehealth apps. In 10 papers, patients were able to use telehealth despite declining or poor health conditions. The apps were perceived as simple, clear, easy, effortless, and not too time consuming to use. Patients felt comfortable using the technology [23,25,27-34]. Although it was easy to use a digital pen and diary, because of the amount of information the patients received about the equipment, the diary, assessment, and the reporting of symptoms, the app was perceived as confusing [32,33]. App simplicity was seen as a prerequisite, especially for older people, in using the technology [22].

In four papers, patients were unable to use the apps because of their poor health condition, physical limitations, or unfamiliarity with the telehealth equipment, and they required assistance from their family [24,25,34,35]. Patients experienced challenges related to the design of the equipment, such as small font size on smartphones, the small size of videophones, or the lack of equipment portability [25,26,34,36]. The required use of a desktop device prevented teleconsultations for bedbound patients and reminded some of approaching death, which was not the case when the desktop device was replaced with a tablet device [36].

Visual Features that Enhance Communication and Care via Telehealth

Patients experienced that telehealth including video was useful for communication and interaction with health care professionals [26,29,34-36,38-40]. Across these studies, patients noted that the visual features allowed them to see the health care professionals they were interacting with, which provided assurance and comfort as well as enhanced care and nursing assessments. The use of video enabled nonverbal communication, such as body language, facial expressions of happiness or suffering, and other emotions, in addition to the situational context [35,36]. According to van Gurp et al [36], this allowed patients and health care professionals to be immersed in a digital connectedness. Communication via video helped health care professionals to discern how patients felt, even when they tried to maintain a facade and pretend that everything was fine [36,39]. Moreover, a physically distant professional listener provided the freedom and privacy that patients needed to talk about difficult issues; thereafter, they continued as usual [36].

In 1 study in which patients were coached on end-of-life communication using webinars, patients reported that this format lacked interpersonal dynamics such as social presence as well as aural and visual communication cues that were considered important because of the sensitive nature of the topic. Nevertheless, patients stated that having the live video of the discussion facilitator and the ability to see other participants could have made it easier to follow the discussions [31]. In another study, although patients indicated a higher level of readiness to use video technology than home care nurses, patients preferred fewer visits overall and preferred to see the home care nurses in person [37].

Symptom Management and Self-Management Promoted by Telehealth

There were equivocal results of whether the use of telehealth improved burdensome symptoms and quality of life. A study testing the feasibility of SMS and interactive voice response found a significant reduction in mean pain score using the European Organization for Research and Treatment of Cancer Quality Questionnaire, but the study found no change while using the Numeric Rating Scale for pain. Furthermore, there was no change in overall quality of life [27]. Another study, using TapCloud for remote monitoring of symptoms, found improved symptom management for moderate to severe dyspnea, moderate to severe depression, and poor well-being [41]. Two RCTs using teleconsultation found no significant differences between the telehealth group and the control group regarding symptom management and quality of life [37,40]. In contrast, an RCT investigating whether weekly teleconsultations from a hospital-based specialist palliative care team improved patients' symptom burden found significantly higher symptom burden in the intervention group than the control group after 12 weeks [42].

Patients perceived that use of telehealth improved quality of care, enhanced self-management of pain, and contributed to more sincere pain reporting [23,26,28,32,38]. Patients who tended to forget which medication they had taken regarded the medication overview on the app as supportive, and those who took their analgesics based on the clock found the visual and sound reminders useful [23]. Patients perceived that the symptoms included in the app were too general [24], and they wanted the possibility to elaborate on their answers that were primarily related to pain, such as multiple localizations, type of pain, and why or how pain changed [24,33]. Furthermore, patients did not agree on how often the app for symptom monitoring should be used [25]. In 1 study, patients assessed symptoms using the Edmonton Symptom Assessment System, both at the hospital and at home via teleconferencing with a multidisciplinary team [35]. This study suggested that teleconferencing allowed improved symptom control.

Perceptions of Improved Palliative Care at Home

Patients felt that the use of telehealth increased and improved their access to health care professionals at home [22-24,26,28,33,34,36,39,43] and perceived that they had increased access to health care professionals in case of emergency, during the night, or on an as-needed basis [26,39].

Furthermore, patients noted that contact with nurses was the most valuable component of the app [23].

It was important for the patients to know that health care professionals were available for them, were looking after them, and were monitoring them by means of telehealth. This contributed to feelings of being cared for at home, connectedness, relief, tranquility, and enhanced security [24,25,33,36,40]. Symptom monitoring at home facilitated communication of symptoms to hospital-based health care professionals [24,28,33], and remote home follow-up was perceived as less intrusive than a phone call [33,40]. During teleconsultations with their palliative care team and general practitioner, patients experienced concentrated responsiveness and possibilities for direct agreement on the division of responsibilities for future actions. However, patients felt insecure and needed to act as mediators when there was disagreement among health care professionals [43].

Discussion

Principal Findings

This scoping review mapped and assessed published studies on patients' experiences of using telehealth in palliative home care. The results showed that telehealth apps seemed to be feasible for use in palliative care, increased and improved access to health care professionals at home, and enhanced feelings of security and safety. The visual features of telehealth allowed a close connectedness with health care professionals, although there were contradicting results on whether the use of telehealth improved burdensome symptoms and quality of life.

Telehealth apps seem to be feasible for use in palliative care and do not seem to add further burden to most patients. The prerequisites for patients' willingness to use telehealth seem to be app simplicity and telehealth services being perceived as valuable to the patients [44]. However, the results indicate that the design issues of the apps had negative consequences of usability and user friendliness for some patients. Design issues may increase the dependency on help from others or act as a reminder of declining health and approaching death. Moreover, experiences of usability and other concerns related to telehealth may differ among patients with different illnesses [45]. To design and deliver apps and telehealth services that are even more closely aligned with patient needs, participation of users in the design process of technical solutions is crucial [46].

The results indicate that the use of telehealth improved access to health care professionals, although patients remained at their own homes. The use of telehealth seems to support the patient's choice of living at home for as long as possible, which is important for many patients [2]. Patients may feel more comfortable, in control, and in a safer place at home compared with being in hospitals. In addition, the home environment seems more conducive to engaging in meaningful activities and relationships [2,36,47]. Patient choice and autonomy are essential in palliative care [48]. Promoting patients' choice about *where* and *how* health care is provided may contribute to patients retaining stewardship over their lives, which is often compromised by illness [49].

The use of telehealth apps may strengthen the relationship between patients and health care professionals [50]. Our results suggest that in using telehealth, patients experienced a genuine relationship with health care professionals who looked after them and provided care according to their needs. This promoted feelings of being cared for and feeling secure. Trust is an important element of the health care professional-patient relationship, and it has been described as "a belief that good will be taken care of, or an attitude bound to time and space in which one relies with confidence on someone or something, and as willingness to engage in oneself in a relationship with acceptance of that vulnerability may arise" [51]. Trust may be regarded as something health care professionals must earn and work hard to attain. Availability and access to health care professionals and feeling physically and emotionally safe are described as important in this scoping review, in addition to respectful communication. These are some of the prerequisites for trust in the health care professional-patient relationship [51]. Patient awareness and the trust in the fact that health care professionals are watching over them may reduce feelings of loneliness with their health condition [52,53]. However, palliative care nurses have expressed concerns that technical issues with telehealth apps could jeopardize the relationship of trust with patients [54].

The finding that the use of video-based technology enhances communication and care is supported by previous reviews [55,56]. In line with a meta-ethnography on the experience of telehealth in patients with COPD [57], visual features using video images contributed to experiences of closeness, despite the remote contact. Although patients seem to highly value the use of telehealth, many simultaneously underline the importance and value of the physical presence of health care professionals at their home [58,59]. This may be especially important for those with limited social networks or poor social relations [2,57,59]. In addition, it may be more challenging to create a trusting relationship remotely than in person [8], and a caring touch, which patients in need of palliative care may appreciate [2], is impossible with remote contact. Some found that it may not be appropriate to discuss serious diagnoses or end-of-life issues via video because of the lack of physical closeness [60,61], whereas others found that the lack of physical closeness helped them address difficult issues with health care professionals. Consequently, health care professionals need to distinguish the appropriateness of using telehealth for communication and, in turn, individually tailor patient care [56,57].

Contradicting results were found on whether the use of telehealth improved burdensome symptoms and quality of life. Two one-group pre-post studies reported some improvements, whereas two of the RCTs found no significant differences between the groups. This is in line with a previous systematic review [13]. RCTs are considered the gold standard for investigating the effectiveness of interventions, as the design minimizes the risk of bias [62]. However, RCTs in palliative care research are often limited by poor recruitment, small sample size, and attrition, and consequently, quasi-experimental and observational design may be justified when randomization is considered inappropriate [63,64]. In the last RCT [42] included

in this scoping review, a significantly higher symptom burden was reported in the experimental group compared with the controls. The patients in the experimental group may have had a higher awareness of symptoms leading to worsening symptom experience, or their symptoms may have been more precisely registered compared with controls, because of weekly teleconsultations [42]. Another explanation could be the increase in patients' honesty in reporting their symptoms when using telehealth [32]. Nevertheless, this may also suggest that improving the symptom burden in these patients is complex.

A theoretical framework for understanding the mechanism of an intervention is recommended when conducting palliative care research on complex interventions [63]. Therefore, it was surprising that none of the studies that investigated whether the use of telehealth improved symptoms or quality of life applied such a framework. The use of theory in the development phase of the intervention is imperative to be able to explain eventual achieved effects [65]. The use of theory seems to be associated with positive results and large effect sizes [66].

Patients' experiences of using telehealth in palliative home care have mostly been studied in populations comprising patients with cancer. Although the origins and the development of palliative care are closely linked to oncology, early integration of palliative care is increasingly emphasized [67,68]. Patients with life-limiting illnesses other than cancer also experience various problems and care needs early in the illness trajectories [69,70]. However, these patients may have a more unpredictable illness progression than patients with cancer, which presents a challenge in identifying the optimal time for introducing palliative care [71,72] and telehealth. However, patients with a life-limiting illness other than cancer may have been included in other studies, without the interventions being classified as palliative home care interventions.

Notably, few studies included the oldest-old patients, although this population increases continuously and also lives longer with life-limiting illness because of improvement in treatments [61,73]. There may be challenges to including the oldest-old patients in telehealth research. The oldest-old patients may not perceive telehealth as an appealing form of interaction with health care professionals, or health care professionals may inadvertently act as gatekeepers believing that old age and rapidly deteriorating health conditions make participation in telehealth research unfeasible [73].

This scoping review indicates that patients' experiences of telehealth in palliative home care has mostly been studied in populations comprising patients with cancer, and few papers included the oldest-old patients. Mixed method is most frequently used for study design, whereas a limited number of papers used an RCT design. Furthermore, none of the papers that investigated whether the use of telehealth improved symptoms or quality of life applied a theoretical framework for their intervention.

A strength of this review was that we used an acknowledged framework for conducting scoping reviews, in addition to the PRISMA-ScR for guiding the reporting of the review. We performed a broad comprehensive and systematic search to identify published studies. Furthermore, the study selection process and data extraction were conducted independently by pairs of authors.

Considering the limitations of this review, different terms and synonyms are used for telehealth and palliative care in the literature [11,74]. There may be terms that we have not been able to identify and include in our search strategy. Patients with life-limiting illnesses other than cancer may have been included in studies without the intervention being classified as a palliative care intervention. Finally, our search strategy had language restrictions as we only included studies in English, Nordic, Spanish, and Portuguese. Owing to these choices, the results may be affected by information bias. Furthermore, only 6 of the 22 included papers were recent papers. Potential sources for heterogeneity in our scoping review are different study populations, diverse use of technologies, and different study designs across the included papers. Consequently, the results related to the patients' experiences of telehealth should be interpreted with caution.

Conclusions

The use of telehealth in palliative home care does not seem to add further burden to most patients. Telehealth increased and improved access to health care professionals at home, and it enhanced the feelings of security and safety. Furthermore, the visual features of telehealth allowed a close connectedness with health care professionals, which seemed to be highly valued. There were contradicting results on whether the use of telehealth improved burdensome symptoms and quality of life. The results further suggest that telehealth apps may be a positive addition to palliative home care, and patients' reports thereof are in favor. However, health care professionals need to individually tailor the telehealth app to enhance usability and user friendliness for patients. Technology including video was preferable to patients.

To make solid inferences and suggest recommendations for practice and policy, more systematic reviews and studies highlighting the negative aspects of telehealth should be conducted. Future studies also need to address the experiences of using telehealth among patients with life-limiting illnesses other than cancer and the oldest-old patients. It is important to investigate whether other populations have different experiences of usability or other concerns regarding telehealth, as compared with patients with cancer and younger patients. Furthermore, studies including RCTs, when appropriate, are required to increase the body of knowledge regarding the effectiveness of telehealth on symptoms and quality of life. The involvement of users in the development of apps and studies is imperative. Using theoretical frameworks to better understand the mechanisms of interventions is important for future knowledge translation and application.

Acknowledgments

The authors would like to acknowledge Kari Larsen Mariussen for helping us to build the search strategy.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy used in Medical Literature Analysis and Retrieval System Online.
[DOCX File , 15 KB - jmir_v22i5e16218_app1.docx]

Multimedia Appendix 2
Characteristics of the included studies.
[DOCX File , 29 KB - jmir_v22i5e16218_app2.docx]

References

1. Radbruch L, Payne S. White Paper on standards and norms for hospice and palliative care in Europe: part 2. Eur J Palliat Care 2010;17(1):22-33 [FREE Full text]
2. Sandsdalen T, Hov R, Høye S, Rystedt I, Wilde-Larsson B. Patients' preferences in palliative care: a systematic mixed studies review. Palliat Med 2015 May;29(5):399-419. [doi: 10.1177/0269216314557882] [Medline: 25680380]
3. Klarare A, Rasmussen BH, Fossum B, Fürst CJ, Hansson J, Hagelin CL. Experiences of security and continuity of care: patients' and families' narratives about the work of specialized palliative home care teams. Palliat Support Care 2017 Apr;15(2):181-189. [doi: 10.1017/S1478951516000547] [Medline: 27443410]
4. Ventura AD, Burney S, Brooker J, Fletcher J, Ricciardelli L. Home-based palliative care: a systematic literature review of the self-reported unmet needs of patients and carers. Palliat Med 2014 May;28(5):391-402. [doi: 10.1177/0269216313511141] [Medline: 24292156]
5. Worth A, Boyd K, Kendall M, Heaney D, Macleod U, Cormie P, et al. Out-of-hours palliative care: a qualitative study of cancer patients, carers and professionals. Br J Gen Pract 2006 Jan;56(522):6-13 [FREE Full text] [Medline: 16438809]
6. Mehta A, Chan LS, Cohen SR. Flying blind: sources of distress for family caregivers of palliative cancer patients managing pain at home. J Psychosoc Oncol 2014;32(1):94-111. [doi: 10.1080/07347332.2013.856057] [Medline: 24428253]
7. Klarare A, Hagelin CL, Fürst CJ, Fossum B. Team interactions in specialized palliative care teams: a qualitative study. J Palliat Med 2013 Sep;16(9):1062-1069. [doi: 10.1089/jpm.2012.0622] [Medline: 24041291]
8. Dorsey ER, Topol EJ. State of Telehealth. N Engl J Med 2016 Oct 6;375(14):1400. [doi: 10.1056/NEJMc1610233] [Medline: 27705257]
9. Busey JC, Michael P. Telehealth--opportunities and pitfalls. J Am Diet Assoc 2008 Aug;108(8):1296-1301. [doi: 10.1016/j.jada.2008.05.002] [Medline: 18656568]
10. Dellifraine JL, Dansky KH. Home-based telehealth: a review and meta-analysis. J Telemed Telecare 2008;14(2):62-66. [doi: 10.1258/jtt.2007.070709] [Medline: 18348749]
11. Stowe S, Harding S. Telecare, telehealth and telemedicine. Eur Geriatr Med 2010;1(3):193-197. [doi: 10.1016/j.eurger.2010.04.002]
12. Tuckson RV, Edmunds M, Hodgkins ML. Telehealth. N Engl J Med 2017 Oct 19;377(16):1585-1592. [doi: 10.1056/NEJMsr1503323] [Medline: 29045204]
13. Head BA, Schapmire TJ, Zheng Y. Telehealth in palliative care: a systematic review of patient-reported outcomes. J Hosp Palliat Nurs 2017 Apr;19(2):130-139. [doi: 10.1097/NJH.0000000000000319]
14. Bradford N, Armfield NR, Young J, Smith AC. The case for home based telehealth in pediatric palliative care: a systematic review. BMC Palliat Care 2013 Feb 1;12:4 [FREE Full text] [doi: 10.1186/1472-684X-12-4] [Medline: 23374676]
15. Zheng Y, Head BA, Schapmire TJ. A systematic review of Telehealth in palliative care: caregiver outcomes. Telemed J E Health 2016 Apr;22(4):288-294. [doi: 10.1089/tmj.2015.0090] [Medline: 26360181]
16. Capurro D, Ganzinger M, Perez-Lu J, Knaup P. Effectiveness of eHealth interventions and information needs in palliative care: a systematic literature review. J Med Internet Res 2014 Mar 7;16(3):e72 [FREE Full text] [doi: 10.2196/jmir.2812] [Medline: 24610324]
17. Johnston B. UK telehealth initiatives in palliative care: a review. Int J Palliat Nurs 2011 Jun;17(6):301-308. [doi: 10.12968/ijpn.2011.17.6.301] [Medline: 21727889]
18. Kidd L, Cayless S, Johnston B, Wengstrom Y. Telehealth in palliative care in the UK: a review of the evidence. J Telemed Telecare 2010;16(7):394-402. [doi: 10.1258/jtt.2010.091108] [Medline: 20813893]
19. Allsop MJ, Powell RA, Namisango E. The state of mHealth development and use by palliative care services in sub-Saharan Africa: a systematic review of the literature. BMJ Support Palliat Care 2018 Jun;8(2):155-163. [doi: 10.1136/bmjspcare-2015-001034] [Medline: 27207721]

20. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005;8(1):19-32. [doi: 10.1080/1364557032000119616]
21. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
22. Aoki N, Ohta S, Yamamoto H, Kikuchi N, Dunn K. Triangulation analysis of tele-palliative care implementation in a rural community area in Japan. Telemed J E Health 2006 Dec;12(6):655-662. [doi: 10.1089/tmj.2006.12.655] [Medline: 17250487]
23. Hochstenbach LM, Zwakhalen SM, Courtens AM, van Kleef M, de Witte LP. Feasibility of a mobile and web-based intervention to support self-management in outpatients with cancer pain. Eur J Oncol Nurs 2016 Aug;23:97-105 [FREE Full text] [doi: 10.1016/j.ejon.2016.03.009] [Medline: 27456381]
24. McCall K, Keen J, Farrer K, Maguire R, McCann L, Johnston B, et al. Perceptions of the use of a remote monitoring system in patients receiving palliative care at home. Int J Palliat Nurs 2008 Sep;14(9):426-431. [doi: 10.12968/ijpn.2008.14.9.31121] [Medline: 19060793]
25. Pinto S, Almeida F, Caldeira S, Martins JC. The Comfort app prototype: introducing a web-based application for monitoring comfort in palliative care. Int J Palliat Nurs 2017 Sep 2;23(9):420-431. [doi: 10.12968/ijpn.2017.23.9.420] [Medline: 28933997]
26. Whitten P, Doolittle G, Mackert M. Telehospice in Michigan: use and patient acceptance. Am J Hosp Palliat Care 2004;21(3):191-195. [doi: 10.1177/104990910402100307] [Medline: 15188918]
27. Besse KT, Boveldt ND, Janssen GH, Vernooij-Dassen M, Vissers KC, Engels Y. Pain assessment with short message service and interactive voice response in outpatients with cancer and pain: a feasibility study. Pain Pract 2016 Mar;16(3):320-326. [doi: 10.1111/papr.12278] [Medline: 25581306]
28. Lind L, Karlsson D. A system for symptom assessment in advanced palliative home healthcare using digital pens. Med Inform Internet Med 2004;29(3-4):199-210. [doi: 10.1080/14639230400005966] [Medline: 15742987]
29. Passik SD, Kirsh KL, Leibee S, Kaplan LS, Love C, Napier E, et al. A feasibility study of dignity psychotherapy delivered via telemedicine. Palliat Support Care 2004 Jun;2(2):149-155. [doi: 10.1017/s1478951504040209] [Medline: 16594244]
30. Tieman JJ, Swetenham K, Morgan DD, To TH, Currow DC. Using telehealth to support end of life care in the community: a feasibility study. BMC Palliat Care 2016 Nov 17;15(1):94 [FREE Full text] [doi: 10.1186/s12904-016-0167-7] [Medline: 27855681]
31. Reinke LF, Griffith RG, Wolpin S, Donesky-Cuenco D, Carrieri-Kohlman V, Nguyen HQ. Feasibility of a webinar for coaching patients with chronic obstructive pulmonary disease on end-of-life communication. Am J Hosp Palliat Care 2011 May;28(3):147-152 [FREE Full text] [doi: 10.1177/1049909110376807] [Medline: 20834035]
32. Lind L. Evaluation of the use of digital pens for pain assessment in palliative home healthcare. Stud Health Technol Inform 2008;136:101-106. [Medline: 18487715]
33. Lind L, Karlsson D, Fridlund B. Patients' use of digital pens for pain assessment in advanced palliative home healthcare. Int J Med Inform 2008 Feb;77(2):129-136. [doi: 10.1016/j.ijmedinf.2007.01.013] [Medline: 17363323]
34. Stern A, Valaitis R, Weir R, Jadad AR. Use of home telehealth in palliative cancer care: a case study. J Telemed Telecare 2012 Jul;18(5):297-300. [doi: 10.1258/jtt.2012.111201] [Medline: 22790013]
35. Hennemann-Krause L, Lopes AJ, Araújo JA, Petersen EM, Nunes RA. The assessment of telemedicine to support outpatient palliative care in advanced cancer. Palliat Support Care 2015 Aug;13(4):1025-1030. [doi: 10.1017/S147895151400100X] [Medline: 25159308]
36. van Gurp J, van Selm M, Vissers K, van Leeuwen E, Hasselaar J. How outpatient palliative care teleconsultation facilitates empathic patient-professional relationships: a qualitative study. PLoS One 2015;10(4):e0124387 [FREE Full text] [doi: 10.1371/journal.pone.0124387] [Medline: 25902263]
37. Hebert MA, Brant R, Hailey D, van der Pol M. Potential and readiness for video-visits in rural palliative homecare: results of a multi-method study in Canada. J Telemed Telecare 2016;12(3_suppl):43-45. [doi: 10.1258/135763306779380093]
38. Miyazaki M, Stuart M, Liu L, Tell S, Stewart M. Use of ISDN video-phones for clients receiving palliative and antenatal home care. J Telemed Telecare 2003;9(2):72-77. [doi: 10.1258/135763303321327911] [Medline: 12699575]
39. Whitten P, Doolittle G, Hellmich S. Telehospice: using telecommunication technology for terminally ill patients. J Comput Mediat Commun 2001;6(4):-. [doi: 10.1111/j.1083-6101.2001.tb00128.x]
40. Wilkinson OM, Duncan-Skingle F, Pryor JA, Hodson ME. A feasibility study of home telemedicine for patients with cystic fibrosis awaiting transplantation. J Telemed Telecare 2008;14(4):182-185. [doi: 10.1258/jtt.2008.070107] [Medline: 18534951]
41. Bonsignore L, Bloom N, Steinhauser K, Nichols R, Allen T, Twaddle M, et al. Evaluating the feasibility and acceptability of a Telehealth program in a rural palliative care population: TapCloud for palliative care. J Pain Symptom Manage 2018 Jul;56(1):7-14. [doi: 10.1016/j.jpainsymman.2018.03.013] [Medline: 29551433]
42. Hoek PD, Schers HJ, Bronkhorst EM, Vissers KC, Hasselaar JG. The effect of weekly specialist palliative care teleconsultations in patients with advanced cancer -a randomized clinical trial. BMC Med 2017 Jun 19;15(1):119 [FREE Full text] [doi: 10.1186/s12916-017-0866-9] [Medline: 28625164]

43. van Gurp J, van Selm M, van Leeuwen E, Vissers K, Hasselaar J. Teleconsultation for integrated palliative care at home: a qualitative study. Palliat Med 2016 Mar;30(3):257-269. [doi: 10.1177/0269216315598068] [Medline: 26269323]
44. Johnston B, Kidd L, Wengstrom Y, Kearney N. An evaluation of the use of Telehealth within palliative care settings across Scotland. Palliat Med 2012 Mar;26(2):152-161. [doi: 10.1177/0269216311398698] [Medline: 21378067]
45. Kaplan B, Litewka S. Ethical challenges of telemedicine and telehealth. Camb Q Healthc Ethics 2008;17(4):401-416. [doi: 10.1017/S0963180108080535] [Medline: 18724880]
46. Clemensen J, Rothmann MJ, Smith AC, Caffery LJ, Danbjorg DB. Participatory design methods in telemedicine research. J Telemed Telecare 2017 Oct;23(9):780-785. [doi: 10.1177/1357633X16686747] [Medline: 28027678]
47. Collier A, Phillips JL, Iedema R. The meaning of home at the end of life: a video-reflexive ethnography study. Palliat Med 2015 Sep;29(8):695-702. [doi: 10.1177/0269216315575677] [Medline: 25805739]
48. Wilson F, Ingleton C, Gott M, Gardiner C. Autonomy and choice in palliative care: time for a new model? J Adv Nurs 2014 May;70(5):1020-1029. [doi: 10.1111/jan.12267] [Medline: 24118083]
49. Wheatley VJ, Baker JI. 'Please, I want to go home': ethical issues raised when considering choice of place of care in palliative care. Postgrad Med J 2007 Oct;83(984):643-648 [FREE Full text] [doi: 10.1136/pgmj.2007.058487] [Medline: 17916873]
50. Rubeis G, Schochow M, Steger F. Patient autonomy and quality of care in telehealthcare. Sci Eng Ethics 2018 Feb;24(1):93-107. [doi: 10.1007/s11948-017-9885-3] [Medline: 28281148]
51. Dinç L, Gastmans C. Trust in nurse-patient relationships: a literature review. Nurs Ethics 2013 Aug;20(5):501-516. [doi: 10.1177/0969733012468463] [Medline: 23426234]
52. Hansen LA, Almqvist F, Ørjasæter N, Kistorp KM. Velferdsteknologi i sentrum (VIS)-evaluering av velferdsteknologi fra et tjenestedesignperspektiv (Evaluation of telehealth from the users' perspective). Tidsskrift for omsorgsforskning (Journal of Caring Science) 2017;3(12):144-152. [doi: 10.18261/issn.2387-5984-2017-02-12]
53. Sand L, Strang P. Existential loneliness in a palliative home care setting. J Palliat Med 2006 Dec;9(6):1376-1387. [doi: 10.1089/jpm.2006.9.1376] [Medline: 17187546]
54. Collier A, Morgan DD, Swetenham K, To TH, Currow DC, Tieman JJ. Implementation of a pilot telehealth programme in community palliative care: A qualitative study of clinicians' perspectives. Palliat Med 2016 Apr;30(4):409-417. [doi: 10.1177/0269216315600113] [Medline: 26290500]
55. Kruse CS, Krowski N, Rodriguez B, Tran L, Vela J, Brooks M. Telehealth and patient satisfaction: a systematic review and narrative analysis. BMJ Open 2017 Aug 3;7(8):e016242 [FREE Full text] [doi: 10.1136/bmjopen-2017-016242] [Medline: 28775188]
56. Lindberg B, Nilsson C, Zotterman D, Söderberg S, Skär L. Using information and communication technology in home care for communication between patients, family members, and healthcare professionals: a systematic review. Int J Telemed Appl 2013;2013:461829 [FREE Full text] [doi: 10.1155/2013/461829] [Medline: 23690763]
57. Barken TL, Söderhamn U, Thygesen E. A sense of belonging: a meta-ethnography of the experience of patients with chronic obstructive pulmonary disease receiving care through telemedicine. J Adv Nurs 2019 Dec;75(12):3219-3230. [doi: 10.1111/jan.14117] [Medline: 31225664]
58. Gorst SL, Armitage CJ, Brownsell S, Hawley MS. Home telehealth uptake and continued use among heart failure and chronic obstructive pulmonary disease patients: a systematic review. Ann Behav Med 2014 Dec;48(3):323-336 [FREE Full text] [doi: 10.1007/s12160-014-9607-x] [Medline: 24763972]
59. Rykkje L, Hjorth G. 'Safety at Home': experiences from testing of video communication between patients and home health care personnel. SAGE Open 2017;7(4):215824401774490. [doi: 10.1177/21582440177449000]
60. Young JD, Borgetti SA, Clapham PJ. Telehealth: exploring the ethical issues. DePaul J Health Care L 2018;19(3):1-15 [FREE Full text]
61. Botrugno C. Towards an ethics for telehealth. Nurs Ethics 2019 Mar;26(2):357-367. [doi: 10.1177/0969733017705004] [Medline: 28502219]
62. Faber T, Ravaud P, Riveros C, Perrodeau E, Dechartres A. Meta-analyses including non-randomized studies of therapeutic interventions: a methodological review. BMC Med Res Methodol 2016 Mar 22;16:35 [FREE Full text] [doi: 10.1186/s12874-016-0136-0] [Medline: 27004721]
63. Higginson IJ, Evans CJ, Grande G, Preston N, Morgan M, McCrone P, MORECare. Evaluating complex interventions in end of life care: the MORECare statement on good practice generated by a synthesis of transparent expert consultations and systematic reviews. BMC Med 2013 Apr 24;11:111 [FREE Full text] [doi: 10.1186/1741-7015-11-111] [Medline: 23618406]
64. Greenhalgh T, Annandale E, Ashcroft R, Barlow J, Black N, Bleakley A, et al. An open letter to The BMJ editors on qualitative research. Br Med J 2016 Feb 10;352:i563. [doi: 10.1136/bmj.i563] [Medline: 26865572]
65. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. Int J Nurs Stud 2013 May;50(5):587-592. [doi: 10.1016/j.ijnurstu.2012.09.010] [Medline: 23159157]
66. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]

67. Kaasa S, Loge JH, Aapro M, Albreht T, Anderson R, Bruera E, et al. Integration of oncology and palliative care: a Lancet Oncology Commission. Lancet Oncol 2018 Nov;19(11):e588-e653. [doi: 10.1016/S1470-2045(18)30415-7] [Medline: 30344075]
68. Evans JM, Mackinnon M, Pereira J, Earle CC, Gagnon B, Arthurs E, et al. Integrating early palliative care into routine practice for patients with cancer: a mixed methods evaluation of the INTEGRATE Project. Psychooncology 2019 Jun;28(6):1261-1268. [doi: 10.1002/pon.5076] [Medline: 30946500]
69. Beernaert K, Deliens L, de Vleminck A, Devroey D, Pardon K, Block LV, et al. Is there a need for early palliative care in patients with life-limiting illnesses? Interview study with patients about experienced care needs from diagnosis onward. Am J Hosp Palliat Care 2016 Jun;33(5):489-497. [doi: 10.1177/1049909115577352] [Medline: 25852203]
70. Kaufman BD, Cohen HJ. Palliative care in pediatric heart failure and transplantation. Curr Opin Pediatr 2019 Oct;31(5):611-616. [doi: 10.1097/MOP.0000000000000799] [Medline: 31335747]
71. Dalgaard KM, Bergenholtz H, Nielsen ME, Timm H. Early integration of palliative care in hospitals: a systematic review on methods, barriers, and outcome. Palliat Support Care 2014 Dec;12(6):495-513. [doi: 10.1017/S1478951513001338] [Medline: 24621947]
72. Iyer AS, Dionne-Odom JN, Ford SM, Tims SL, Sockwell ED, Ivankova NV, et al. A formative evaluation of patient and family caregiver perspectives on early palliative care in chronic obstructive pulmonary disease across disease severity. Ann Am Thorac Soc 2019 Aug;16(8):1024-1033. [doi: 10.1513/AnnalsATS.201902-112OC] [Medline: 31039003]
73. Worster B, Swartz K. Telemedicine and palliative care: an increasing role in supportive oncology. Curr Oncol Rep 2017 Jun;19(6):37. [doi: 10.1007/s11912-017-0600-y] [Medline: 28417310]
74. Hui D, Mori M, Parsons HA, Kim SH, Li Z, Damani S, et al. The lack of standard definitions in the supportive and palliative oncology literature. J Pain Symptom Manage 2012 Mar;43(3):582-592 [FREE Full text] [doi: 10.1016/j.jpainsymman.2011.04.016] [Medline: 22104619]

Abbreviations

CINAHL: Cumulative Index to Nursing and Allied Health
COPD: chronic obstructive pulmonary disease
MEDLINE: Medical Literature Analysis and Retrieval System Online
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 11.09.19; peer-reviewed by C Kruse, J Meirte; comments to author 17.12.19; revised version received 18.01.20; accepted 10.02.20; published 05.05.20.

Please cite as:
Steindal SA, Nes AAG, Godskesen TE, Dihle A, Lind S, Winger A, Klarare A
Patients' Experiences of Telehealth in Palliative Home Care: Scoping Review
J Med Internet Res 2020;22(5):e16218
URL: https://www.jmir.org/2020/5/e16218
doi:10.2196/16218
PMID:32369037

Review

Sexual Behaviors and Violence in Pornography: Systematic Review and Narrative Synthesis of Video Content Analyses

Elise R Carrotte[1], BPsych (Hons), MPsych; Angela C Davis[1,2], PhD, BA (Hons); Megan SC Lim[1,2,3], PhD, BBiomedSci (Hons)

[1]Burnet Institute, Melbourne, Australia

[2]Monash University, Melbourne, Australia

[3]University of Melbourne, Melbourne, Australia

Corresponding Author:
Megan SC Lim, PhD, BBiomedSci (Hons)
Burnet Institute
85 Commercial Rd
Melbourne, 3004
Australia
Phone: 61 0385062403
Email: megan.lim@burnet.edu.au

Abstract

Background: Owing to increasing access to Web-based pornography and concerns about its impact on viewers, many researchers have attempted to systematically analyze the content of pornography.

Objective: We aimed to systematically review the results of quantitative content analyses of video-based pornography and identify the degree to which the following behaviors are depicted: (1) sexual behaviors and themes, (2) condom use during sexual behaviors, and (3) aggression and violence.

Methods: Inclusion criteria for article eligibility were (1) peer-reviewed publications, (2) articles in the English language, (3) articles describing a quantitative content analysis of video pornography, and (4) articles quantitatively examining at least one variable of interest.

Results: A total of 23 studies met the eligibility criteria. Studies varied in scope and definitions of behavioral variables. Condom use was rare, although more commonly depicted in gay male pornography (36%-64% videos) compared with heterosexual pornography (2%-3% videos). Normative sexual behaviors were most frequently depicted in pornography samples (eg, vaginal intercourse in 48%-90% and fellatio in 52%-90% of heterosexual videos; fellatio in 66%-100% and anal intercourse in 70%-80% of gay male videos). Extreme acts of violence (1%-3% videos) and rape (0%-6% videos) were relatively rare. However, more subtle forms of aggression, such as spanking (5%-75% videos), were more common, and unequal sexual relations (eg, domination) were also common. Although estimates varied by study, dominating and violent behaviors were nearly always directed toward women.

Conclusions: Condom nonuse and gender inequalities are common in pornography, which has implications for the development of healthy sexual relationships among pornography viewers. Higher quality research, including study replication and consistent methodological choices, is needed.

(J Med Internet Res 2020;22(5):e16702) doi:10.2196/16702

KEYWORDS

pornography; content analysis; sexual media; sexual violence; sexual behavior

Introduction

Background

Pornography, in various forms, has existed throughout human history. Different concerns about the content of pornography have periodically emerged over history as social and sexual norms shift. For example, some earlier concerns related to the controversy of depicting nudity and sexual acts in an uncensored manner, whereas more recently concerns in some areas have shifted to “extreme” or “hard-core” violence and its impact on the viewer. These concerns have driven studies of pornographic content and informed their analytical framework.

In the internet age, pornography use is more popular and accessible than ever before; for example, in one survey of young people aged 15 to 29 years, 100% of young men and 81% of young women had ever viewed pornography [1]. With consideration of its nearly ubiquitous nature, pornography is argued to influence the sexual socialization of its viewers, impacting upon beliefs of what is normal and desirable in sexual contexts [2]. Young people have reported that viewing pornography can allow exploration and education around sexuality via exposure to new and appealing behaviors [3,4]. Regardless, it is most likely that pornography's effect on the viewer, be it positive or negative, are at least partially impacted by the types of behaviors that they are viewing [2].

Research has found links between pornography use and potential harms. For example, a large body of research indicates that pornography use is associated with greater unsafe sexual practices and more sexual partners [5,6], and condomless sex among men who have sex with men [7]. Around half of men who have sex with men have reported perceptions that pornography has contributed to their engagement in "riskier" sexual behaviors, while over 90% have fantasized about engaging in similar acts to those seen in pornography [7]. Young people have reported a perception that pornography impacts their sexual relationships, including influencing their understandings of what is normal and desirable, leading to pressure from partners to perform certain acts (eg, young women have reported feeling pressured to engage in anal intercourse with their male partners [8]).

Many individuals have expressed concerns that with the rise of the internet, pornography is becoming more extreme as well as more accessible. Estimates indicate that while 18% of adolescents have been exposed to affection-themed pornography, 18% have been exposed to pornography with themes of dominance, and 10% have been exposed to violent pornography, with exposure to more extreme genres increasing with age [9]. With consideration of the abundance of free pornography currently available, pornography producers have described a competitive climate in which films must show increasingly extreme or niche acts across multiple mediums in order to stand out from other pornography films and appeal to as many sexual desires as possible [10]. In qualitative research, young people have described acts that some people would describe as degrading or violent as being common and normalized in pornography, such as ejaculating on women's faces and apparently nonconsensual bondage, dominance, and sadomasochism (BDSM) [11,12]. Frequency of pornography use and the number of pornography genres viewed have both been associated with higher sexual preferences for types of sexual practices presented in pornography [6]. Of particular concern, a recent meta-analysis has found that using pornography is associated with increased sexual aggression for both men and women, with stronger associations for verbal than physical aggression [13]. The relationship between sexual aggression perpetration and intentional exposure to pornographic material appear to be stronger for violent pornography compared with nonviolent pornography [14]. Pornography use has also been associated with stronger attitudes supporting violence against women [15].

In response to these arguments, several studies have attempted to quantify what pornography depicts in order to determine the messages viewers may be internalizing. These content analyses, which aim to be systematic and replicable analyses of media [16], have ranged in scope, methodology, and complexity. This means that there are various results which can inform policy, but also risk being misinterpreted. For example, one content analytic study [17] reported that 88% of pornographic scenes depicted physical aggression in their analysis; this statistic is frequently used in articles and opinion pieces arguing that pornography is violent and harmful [18,19]. However, other studies have found much smaller estimates of physical violence, such as 13% of internet videos [20] or 1% of video and DVD scenes [21]. This relates to differences in sampling methodology (eg, sampling "popular" pornography, subgenres, or different mediums) and issues with behavioral definitions (eg, only coding aggressive acts that researchers interpret as being nonconsensual [22]). Therefore, there is a need to systematically synthesize content analyses of pornography to provide more accurate estimates of behaviors depicted in pornography.

Objectives

This systematic review aimed to review and synthesize the results of content analysis study designs that quantitatively analyzed content of video-based pornography. The outcomes measured were the frequency with which the following types of behaviors are depicted: (1) sexual behaviors and themes, (2) condom use, and (3) aggression and violence.

Methods

Study Design

This systematic review was conducted in accordance with Cochrane methodologies [23]. Methods and results are presented in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement [24]. A protocol was developed by the authors prior to the study to guide the search and data extraction; this protocol has not been published.

Eligibility Criteria

Inclusion criteria for article eligibility were (1) peer-reviewed publication in an academic journal; (2) English language; (3) described a quantitative content analysis of video pornography, and (4) quantitatively examined at least one variable of interest (Table 1). A quantitative content analysis was understood to be an analysis "which involves the creation and use of pre-determined categories for the purpose of understanding and describing media messages in a way that can be counted and quantified" [25]. Quantitative content analyses were chosen over qualitative or narrative approaches as they are the most common type of content analysis and the use of systematic categories allows relatively easy comparison [25]. Video pornography is understood here as video-based material containing the explicit display of sexual organs or activity, intended to stimulate sexual excitement, thereby eliminating solely still image, written or audio erotica from the review.

Table 1. Variables included in the systematic review.

Category, variables, and subvariables	Example description[a]
Sexual behaviors	
Kissing	Kissing between actors on mouth
Oral sex	
Fellatio	Oral-penile contact between actors
Cunnilingus	Oral-vulva or oral-vaginal contact between actors
Anilingus	Oral-anal contact (a.k.a. rimming) between actors
Intercourse	
Vaginal intercourse	Penetration of one actor's vagina by another's penis
Anal intercourse	Penetration of one actor's anus by another's penis
ATM	"Ass-to-mouth"; where an actor inserts their penis into the mouth of another actor after engaging in anal intercourse
Bondage, domination, and sadomasochism (BDSM)	
BDSM (any)	Group category–sexual interaction between two or more people involving domination (one person clearly leading the sexual activity with another person clearly following or submitting) and bondage (physical confinement) of one or more actors
Domination and submission	Depiction of dominating or submissive behaviors (without specifying the presence of bondage)
Bondage	Depiction of bondage behaviors (without specifying the presence of domination or submission)
Group sex	
Group sex (any)	Sexual contact between three or more actors, including threesomes or orgies
Vaginal-anal double penetration (vaginal-anal DP)	Simultaneous penetration of one actor's vagina and anus with two actors' penises
Anal double penetration (anal DP)	Simultaneous penetration of one actor's anus with two actors' penises
Paraphilias	
Bestiality	Sexual interaction between an actor and an animal
Incest	Sexual interaction between actors explicitly described or depicted as relatives
Pedophilia	Sexual interaction involving children
Condom use	Any depiction of condom use by actors, by type of sexual behavior
Orgasms	
Any orgasm	Presence of visible ejaculate or other indicators of orgasm, by gender of actor (eg, shuddering and verbal statements communicating orgasm)
Ejaculation location	Ejaculation by one actor onto another (eg, on face or in mouth)
Aggression and violence	
Verbal aggression	
Name calling or insulting	One actor calling another names or slurs, or otherwise insulting verbally
Physical aggression	
Hair pulling	One actor pulling or tugging another's hair
Spanking	One actor slapping another on buttocks (open-handed)
Slapping	One actor open-handed slapping another on other location on body (not buttocks)
Gagging	One actor inserting their penis very far into another actor's mouth; may or may not stimulate gag reflex
Choking	One actor places hands around another actor's neck; may or may not squeeze
Punching	One actor strikes another with closed fist
Kicking	One actor strikes another with foot or feet
Torture	Infliction of severe or excruciating physical pain by one actor to another

Category, variables, and subvariables	Example description[a]
Murder	Killing of an actor by another
Rape	One actor forces another to engage in sexual intercourse in the absence of consent

[a]These are example definitions; definitions were not always provided within studies, or sometimes differed slightly between studies. Author discretion was used when comparing categories across studies based on definitions. Behaviors included within each variable may be real or simulated by actors.

Variables included in this systematic review are described in Table 1. Variables were chosen using first a deductive approach, focusing on sexual behaviors, condom use and various indicators of aggression and violence. These variables were identified a priori in accordance with public health concerns relating to the prevalence of certain sexual scripts and violence against women [26]. After immersion in the literature, further sexual behavior variables (including those relating to domination of one person over another) and behaviors often conceptualized as illegal, deviant, or socially unacceptable (eg, bestiality, pedophilia, and incest) were chosen. The authors of 3 studies were contacted to provide additional data, which was able to be provided by the authors of one study [27].

Search Procedure and Data Sources

A systematic search of 7 databases, covering all years until the search date, was conducted over a one-week period in March 2016. The search was applied to MEDLINE, PsycINFO, Scopus, EMBASE, CINAHL Plus, Communication and Mass Media Complete, and Cochrane Library databases. Past research has identified challenges searching for content analyses as not all studies use the term "content analysis" [25]. To increase the likelihood of finding articles, additional terms were included. Search terms were developed by all 3 authors and included a combination of a term for pornography and a term indicative of a content analysis [25]. For example, the Medline search was *(porn* or "sexually explicit" or erotic*) OR Erotica/ AND (content ADJ2 analy* OR content ADJ2 review*OR "descriptive study" OR descri* ADJ2 content OR evaluat* ADJ2 content OR map* ADJ2 content)*. The search was repeated in September 2017, to identify any newly published articles.

Data Collection Process

Inclusion of articles was determined in three stages. At each stage, if the authors did not agree on the inclusion of an article, it was discussed and the senior author (ML) was consulted if necessary. First, two authors (EC and AD) each screened all titles for eligibility according to inclusion and exclusion criteria. Abstracts for titles deemed relevant were then screened by the two authors. Finally, articles were subjected to full-text review by either EC or AD to confirm eligibility. Finally, the reference lists of studies in the full-text review were screened by EC and AD to identify potential articles that were not found during the database search, using a similar process. Ambiguities were discussed between authors.

Data Extraction and Risk of Bias

Data extracted included: date of publication, pornography medium (ie, video home system [VHS], DVD, or internet-based video), sampling methods, sample size (eg, number of videos in sample), unit of analysis (eg, scene), variables analyzed, and results.

No best-practice tools are available for assessing bias in systematic reviews of content analyses that are known to the authors. A customized risk of bias tool was created by the authors, based on the Johanna Briggs Institute Critical Appraisal Checklist [28] with reference to content analysis best practice recommendations [29-31]. The 16 checklist items included the reporting of clear inclusion and exclusion criteria, appropriate sampling strategies, development of coding procedures, characteristics of coders, use of appropriate measures of interrater reliability, and measurement of outcomes. At least two authors completed the checklist for each article, with results discussed to reach a conclusion. Risk of bias was deemed "low" if "yes" was selected for 13 to 16 items, "medium" if 9 to 12 items, and "high" if 0 to 8 items (ie, less than half of the criteria were met). A copy of this checklist is available from the corresponding author upon request.

Synthesis of Results

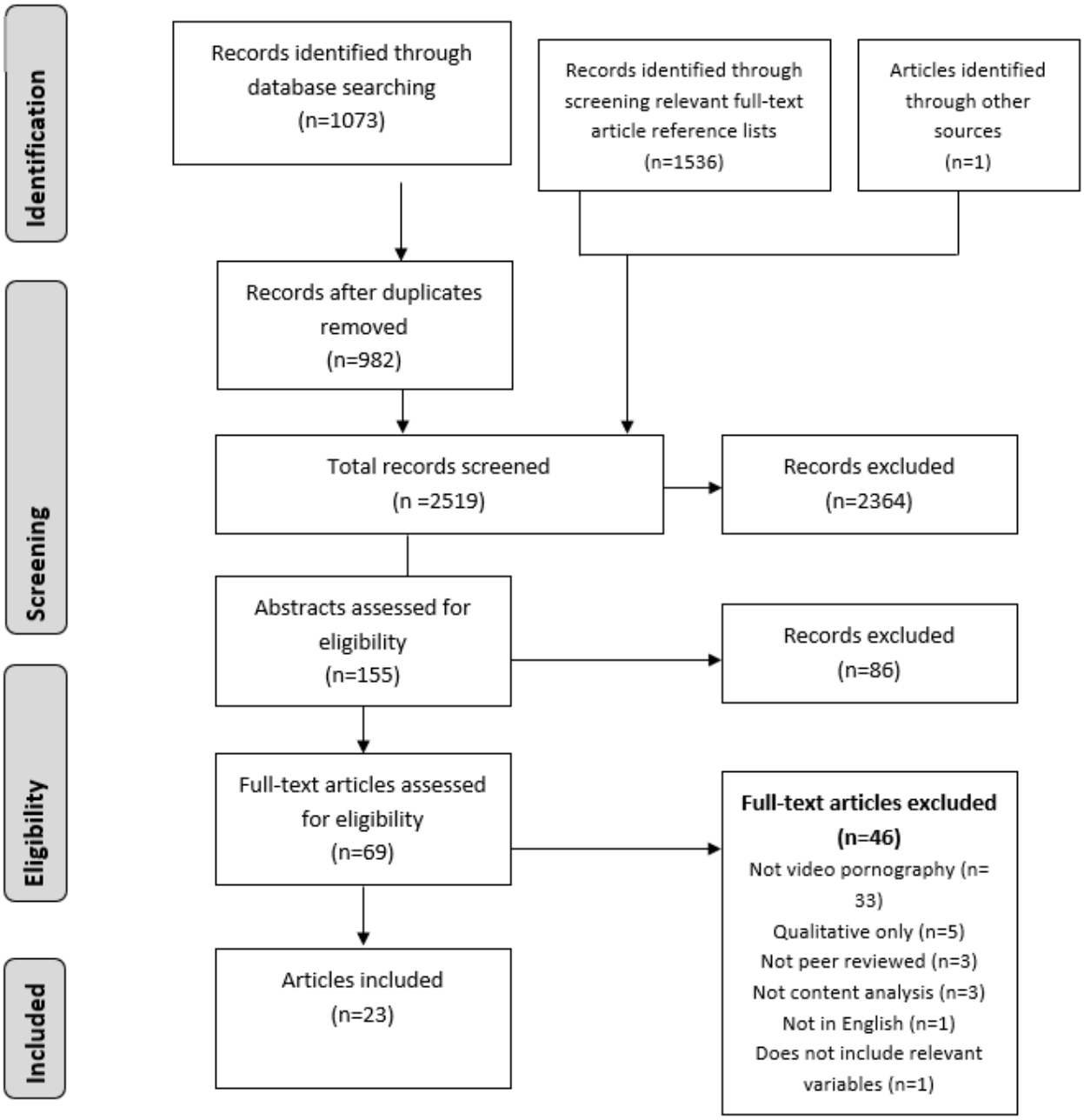
No meta-analyses were conducted due to the heterogeneity of designs, samples, definitions, analyses, and outcome measures. Performing a meta-analysis for such heterogeneous data increases the risk of making erroneous conclusions [32]. Instead, a narrative approach was taken to describe the results [33]. Results were grouped and presented based on the behaviors studied within the sample (eg, sexual behaviors), with consideration of the year of the study, the medium of the pornography (eg, DVD) and the genre (eg, heterosexual). To simplify the presentation of data, frequencies are rounded to the nearest integer, unless less than 1% or greater than 99%. Means are presented rounded to one decimal point. Results are only discussed if providing a specific descriptive statistic for relevant variables (eg, results are not included if the authors described behaviors as simply "rare" or "frequent").

Results

Study Selection

The study selection procedure is illustrated in in Figure 1. The initial database search was conducted in March 2016, producing 1035 articles, and repeated in September 2017, identifying 38 new citations. After removal of duplicates, 982 articles remained. A total of 155 articles were included in the abstract review and 69 articles for the full-text review. In all, 17 studies met the inclusion criteria through database searching. Reference lists of full-text articles were screened, with 1536 articles identified, and 5 articles included. One additional article [34] was located which was not identified through these searches, presumably due to the recent nature of its publication. A total of 23 articles thereby met all the inclusion criteria across both searches.

Figure 1. Screening process and article exclusion.



Sample Characteristics

Table 2 provides a summary of study characteristics. Dates of publication ranged from 1986 to 2017, with the oldest film extracted from 1961 [35] and the most recent in 2015 [34]. Studies varied with regards to the medium of pornography studied, and how data were summarized. Most studies described data using frequencies or percentages (eg, percentage of scenes depicting an act), while some provided means (eg, mean number of acts per video). Studies which provided data by time (eg, amount of time in videos dedicated to an act) are not included in the synthesis due to difficulty comparing with other studies. As seen in Table 2, the most common medium was VHS, followed by internet videos, and DVDs. Most studies were appraised as having a medium risk of bias (11/23, 47.8%), while 8 studies (37.8%) had a low risk of bias and 4 studies had a high risk of bias (17.4%).

Table 2. Characteristics of content analyses identified in the review, by year of publication.

Sr. No.	Year of publication	First author	Content type	Genre (author defined)	Sampling	Unit of analysis	Data type	Behaviors studied relevant to review	Risk of bias
1	1986	Palys [35]	VHS (XXX-rated and "adult" videos)	Unspecified, but intentionally included an "extreme" violent subset	Two videos randomly selected from various video home systems (VHS) rental stores until 125 titles chosen. A further 25 violent videos chosen based on a list of videos that had been charged based on obscenity.	Scenes with either sex or aggression (n=2102)	Frequencies	Fellatio, cunnilingus, vaginal intercourse, anal intercourse, incest, bondage, punching, kicking, torture, rape, murder, domination	Medium
2	1988	Cowan [36]	VHS	Unspecified but predominantly heterosexual ("none seemed targeted at a homosexual population")	Purposive selection 45 "widely available" videos from 7 VHS rental stores.	Sex scenes (n=443)	Frequencies	Ejaculation, incest, name calling, bondage, rape, domination	Low
3	1991	Garcia [37]	VHS	Unspecified	Students in a human sexuality class selected the first video they saw in a VHS rental store.	Videos (n=20)	Means	Fellatio, cunnilingus, vaginal intercourse, anal intercourse, group sex	High
4	1990	Prince [38]	VHS	"Classic" "Popular" films	Purposive selection of popular or well-known feature films produced between 1972 and 1985.	Sex scenes (n=248), actors (n=325), violent acts (n=44)	Frequencies	Ejaculation, orgasms, bestiality, incest, murder	High
5	1990	Yang [39]	VHS (X-rated and XXX-rated)	Unspecified	Randomly selected 30 X- and 30 XXX-rated videos from a VHS rental catalogue.	1639 behavioral sequences (n=984 in X rated, n=655 in XXX rated)	Frequencies	Incest, rape	Medium
6	1991	Duncan [40]	VHS	Unspecified	Randomly selected 50 videos (10% of total available) in a local VHS rental store.	Scenes (n=822)	Frequencies	Incest, bondage, rape	Medium
7	1993	Brosius [41]	VHS	Unspecified but "predominantly heterosexual"	Randomly selected 50 videos from archives of a German copyright firm.	Sex scenes (n=436)	Frequencies	Fellatio, cunnilingus, vaginal intercourse, anal intercourse, group sex, orgasms, domination	Low
8	1994	Cowan [36]	VHS	"Interracial"	Purposive selection of 54 videos in a VHS rental store based on the presence of interracial themes.	Actors (n=476)	Frequencies	Kissing, fellatio, intercourse, cunnilingus, ejaculation, name calling, rape, submission	Medium
9	1999	Monk-Turner [42]	VHS (X and XXX-rated)	Heterosexual	Randomly selected 40 videos from a list of all videos available from a national chain. Only analyzed videos around 2 hours in length and excluded videos with all male or all female actors.	Vignettes (n=209)	Frequencies	Ejaculation, bestiality, name calling, torture, rape, murder	High
10	2000	Barron [43]	VHS	Unspecified	Randomly selected 10 videos in each of 5 stores.	Scene (n=364)	Frequencies	Choking, bondage, punching, kicking, torture, murder, domination	Medium

Sr. No.	Year of publication	First author	Content type	Genre (author defined)	Sampling	Unit of analysis	Data type	Behaviors studied relevant to review	Risk of bias
11	2005	McKee [21]	DVDs and VHS	"Mainstream"	Selected 50 most popular videos across two large video or DVD mail-order companies.	Sex scenes (n=642)	Frequencies, length of scene	Name calling, spanking, slapping, bondage, rape, dominance	Low
12	2008	Sun [44]	DVDs	Unspecified	Sampled 44 videos from a list of most frequently rented VHS and DVDs from a leading trade journal, including 11 female-directed films and 33 male-directed films. Random sample of 61 male-directed scenes selected to match the 61 scenes directed by women.	Scenes (n=61 male directed, n=61 female directed)	Frequencies	Kissing, fellatio, cunnilingus, vaginal intercourse, anal intercourse, double penetration (DP), ass-to-mouth (ATM), group sex, ejaculation, condom use, name calling, spanking, slapping, hair pulling, choking, gagging, bondage, punching, kicking, torture.	Low
13	2009	Grudzen [45]	DVDs	Heterosexual and gay male	Randomly selected DVDs from the largest US distributor of pornographic films, then randomly selected one sex scene from each film.	Sex scenes (n=50 heterosexual, n=50 gay male)	Frequencies	Fellatio, cunnilingus, vaginal intercourse, anal intercourse, anilingus, ATM, group sex, ejaculation, condom use.	Low
14	2010	Bridges [17]	DVD	Unspecified "popular"	Randomly selected 50 videos from the best-selling and most rented lists.	Scenes (n=304), aggression-level acts (n=3375)	Frequencies	Fellatio, cunnilingus, vaginal intercourse, anal intercourse, DP, ATM, group sex, ejaculation, bestiality, condom use, name calling or insulting, spanking, slapping, choking, gagging, bondage, punching, kicking, torture, murder	Low
15	2010	Gorman [20]	Internet videos	Popular free videos (professional and amateur)	Searched Google.com for free pornographic websites, randomly sampled every 5th video from the first five websites per search term.	Videos (n=45)	Frequencies	Kissing, fellatio, cunnilingus, vaginal intercourse, anal intercourse, group sex, ejaculation, condom use, name calling, dominance	Medium
16	2012	Salmon [46]	DVDs	Heterosexual and homosexual best-selling DVDs	30 heterosexual and 30 homosexual DVDs selected from bestselling or top rental lists available through an Web-based adult DVD rental service.	DVD films (n=30 heterosexual, n=30 homosexual)	Means	Fellatio, cunnilingus, vaginal intercourse, anal intercourse, ejaculation, coercion	High
17	2014	Downing [47]	Internet videos	Gay male (amateur and professional) from highly trafficked websites	Selected most recently watched or uploaded video from gay male section of five purposely chosen websites at various times.	Videos (n=302)	Frequencies	Kissing, fellatio, anal intercourse, anilingus, group sex, ejaculation, condom use, spanking, bondage	Low

Sr. No.	Year of publication	First author	Content type	Genre (author defined)	Sampling	Unit of analysis	Data type	Behaviors studied relevant to review	Risk of bias
18	2014	Peters [48]	Internet videos	Popular teen	Identified three popular pornographic websites through Google.com and Alexa.com. Randomly selected 50 videos from "teen" section of each website. Excluded amateur, animated, non-English videos.	Videos (n=150)	Frequencies	Kissing, fellatio, cunnilingus, vaginal intercourse, anal intercourse, anilingus, spanking, bondage, rape	Medium
19	2014	Vannier [49]	Internet videos	Teen compared with mother(s) I'd like to fuck (MILF)	Using Google.com search, identified pornographic websites and excluded those requiring payment, were interactive, or did not have teen and MILF categories. Selected five MILF and five teen videos randomly from each website over a 2-month period.	Videos (total n=100; MILF=50, teen=50).	Frequencies	Kissing, fellatio, cunnilingus, vaginal intercourse, anal intercourse, anilingus, ejaculation, condom use, spanking, domination	Medium
20	2015	Klaassen [27]	Internet videos	"Popular," "mainstream" heterosexual	Selected most 100 viewed videos on four popular pornographic websites and coded first scene of each video. Excluded cartoons, nonsexual videos.	First sex scenes (n=400)	Frequencies	Fellatio, cunnilingus, orgasms, spanking, slapping, pulling hair, choking, gagging, bondage, punching, kicking, torture, rape, murder, domination	Low
21	2016	Zhou [50]	Internet videos	"Asian women" videos compared with other popular categories (eg, teen, MILF, blonde, big tits)	Selected videos from top 10 categories (including "Asian women" category) on xvideos.com, a popular pornographic website. Used systematic, stratified sampling method.	Scenes (n=3132 total, including 172 "Asian women" scenes)	Frequencies	Kissing, fellatio, cunnilingus, anilingus, vaginal intercourse, anal intercourse, ATM	Medium
22	2017	Fritz [51]	Internet videos	"Feminist," "for women" and "mainstream" pornography	Randomly selected videos from Lust Cinema, a website nominated for the Feminist Porn Awards. Also selected content from CrashPad Series, a "queer" feminist site, to diversify selection. Randomly selected videos from "For Women" category on PornHub. Also, randomly selected videos from five largest categories on PornHub to form "mainstream" sample (from categories "teen," "big tits," "brunette," "amateur," "blonde").	Scenes (n=300 total –100 feminist, 100 "for women," 100 "mainstream").	Frequencies	Vaginal intercourse, ejaculation, orgasm, bondage, dominance, and sadomasochism (BDSM), domination	Medium
23	2017	Séguin [34]	Internet videos	Popular	Selected 50 most viewed videos of all time from pornhub.com across all categories.	Videos (n=50) and instances of orgasms (39 male, 20 female).	Frequencies	Orgasms	Medium

Content of interest varied across studies. Types of pornography studied, when specified, were typically heterosexual or "general" or "popular" or "mainstream" videos. Some studies did not specify a genre but noted that their data collected was predominantly aimed at a heterosexual audience. For the purpose of data synthesis, studies are grouped as either "heterosexual" or "gay male." Other analyses specifically studied pornography labeled as interracial [36], "Asian women" [50], feminist/ "for women" [51], teen or youth [48,49], and MILF ("mother(s) I'd like to fuck") [49]. As these genres were likely to be predominantly aimed at a heterosexual audience, these categories were also included in the "heterosexual" group for the purpose of discussion. The authors of this study recognize this is a limitation collapsing broad genres into an overall category and urge readers to seek individual studies for more detailed discussions of their findings. Older studies typically analyzed pornography on a scene by scene basis, whereas internet studies typically analyzed by video, presumably because most internet pornography videos typically involve only one sex scene compared with a feature film or movie interspersed with sex scenes [27].

Sexual Behaviors and Themes

Most studies examined at least one sexual behavior variable. Data for key variables of interest are summarized in Figures 2 and 3 by genre and medium. Where a study presented results by different types or genres of pornography (eg, teen vs MILF [49]), estimates are presented separately here.

Kissing

In all, 7 studies examined kissing. Among heterosexual genre videos, the proportion of scenes depicting kissing ranged from 8% to 50%. Only 1 study specifically examined gay male pornography [47], finding that 34% of Web-based videos included kissing.

Oral Sex

A total of 10 studies specifically examined female to male (F-M) fellatio. The oldest study, an analysis of VHS sex scenes [41], found 54% depicted F-M fellatio. A study of DVDs found that 93% of male-directed scenes involved F-M fellatio, compared with 67% of female-directed scenes [44]. In other pornographic DVD studies, estimates ranged from 84% [45] to 90% [17] of sex scenes, or a mean of 7.0 instances per film [46]. Among 5 studies examining internet videos, F-M fellatio was depicted in 52% to 90% of videos [20,27,48,50].

In all, 2 studies examined male to male (M-M) fellatio in gay male DVDs, with 1 study finding a mean of 11.3 instances per film [46] and another finding 100% of films depicting M-M fellatio [45]. Another study found 66% of gay male internet videos depicted M-M fellatio [47]. Meanwhile, 2 studies examined M-M fellatio in heterosexual or general pornography, with estimates ranging from 0% to 2% of DVD sex scenes [17,44].

Cunnilingus was coded similarly, with estimates differing based on whether the study provided a general overview of cunnilingus or broke behavior down by the gender of actors. The 7 studies that provided a general "cunnilingus" category are presented first. Among VHS studies, cunnilingus was found in 40% to 72% of sex scenes [35,41], or a mean of 4.8 incidents per video [37]. In a DVD study, Salmon and Diamond [46] found a mean of 6.5 instances per heterosexual film, and a mean of zero acts in gay male films. Internet studies found cunnilingus in 41% to 48% of videos [27,48,49].

Some studies broke cunnilingus down by gender. In 2 DVD studies, M-F cunnilingus was found in 43% to 56% of scenes [17,44]. The same 2 studies found that F-F cunnilingus was depicted less frequently [17,44]. Only one study has examined gendered cunnilingus in internet pornography, finding M-F cunnilingus in 18% of scenes and F-F cunnilingus in 6% of scenes [50].

In all, 3 studies specifically examined anilingus, all using internet samples, and 2 studies examined general or heterosexual pornography; 1 study found that 14% of internet videos depicted anilingus [49] while another study found M-F anilingus in 5% scenes and F-F anilingus in 1% scenes [50]. Another study [47] examined internet gay male pornography, finding that 17% of Web-based videos include anilingus.

Figure 2. Percentage of general/heterosexual samples depicting key sexual behaviors, by medium of pornography (VHS, DVD, or internet-based video). Notes: Results are only presented in this graph if the study presented results by percentage of scenes or videos (ie, studies that presented results by time, mean, character excluded from graph). Paraphilias not presented due to sparse data. BDSM and group sex–related behaviors not presented due to variance in definitions and categories between studies. Ejaculation variable is presented in accordance with any estimate of ejaculation; if reporting multiple positions, the highest estimation is provided in this graph. VHS: video home system; MILF: mother(s) I'd like to fuck; BDSM: bondage, dominance, and sadomasochism.

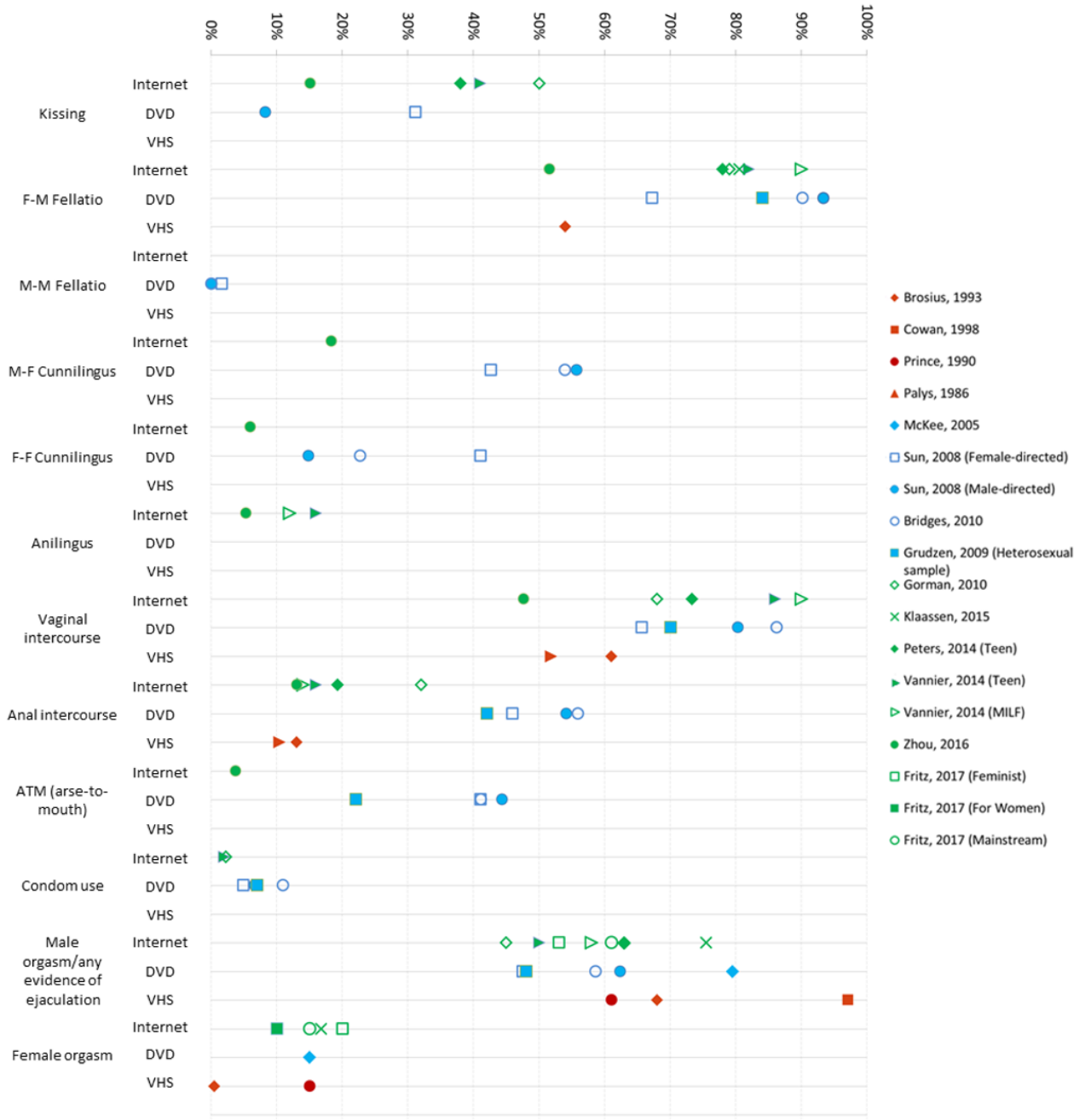
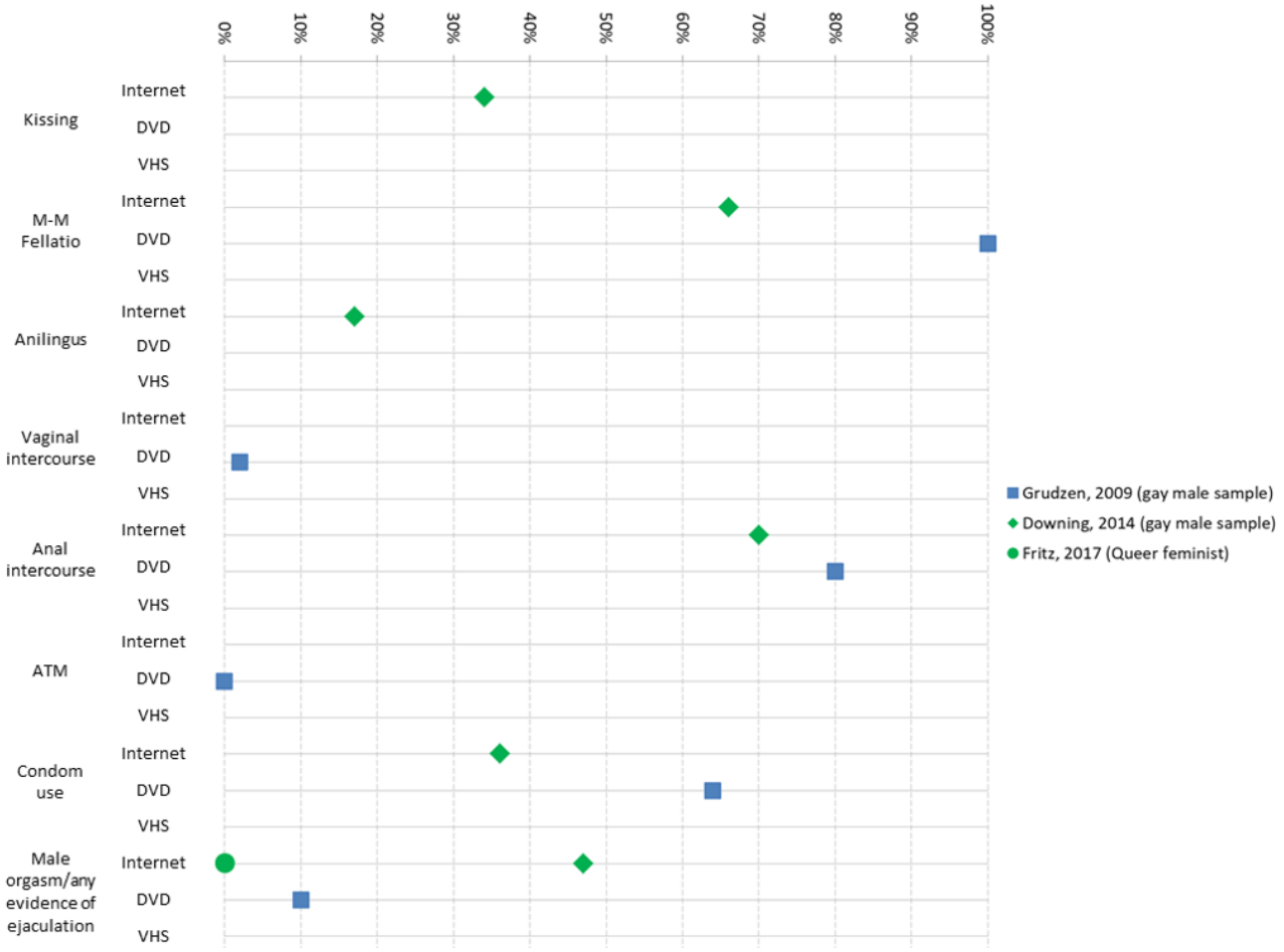


Figure 3. Percentage of gay male/queer samples depicting key sexual behaviors, by medium of pornography (DVD, VHS, or internet-based video), for behaviors studied in these samples. Notes: Results are only presented in this graph if the study presented results by percentage of scenes or videos (ie, studies that presented results by time, mean, character excluded from graph). Paraphilias not presented due to sparse data. Bondage, dominance, and sadomasochism and group sex–related behaviors not presented due to variance in definitions and categories between studies. Ejaculation variable is presented in accordance with any estimate of ejaculation; if reporting multiple positions, the highest estimation is provided in this graph. VHS: video home system.



Vaginal Intercourse

Vaginal intercourse was examined by eleven studies, predominantly in general or heterosexual samples. One VHS study [37] finding a mean of 4.2, 2.4 and 4.2 acts for positions "male above," "female above," and "rear entry" respectively per film. Another found that 31.7% of characters began sexual activity with intercourse but did not provide an overall estimate of its prevalence [36]. One study found vaginal intercourse in 61% of scenes, or 81% of scenes that involved at least one male and one female actor [35,41], while another study found it in 52% [35] of videos. DVD specific studies found vaginal intercourse in 66% to 86% [17,44,45] of sex scenes, while Salmon and Diamond [46] found a mean of 8.5 instances per DVD film. Studies of internet videos have found that vaginal intercourse appears in 48% to 90% [20,48-50] of videos.

Anal Intercourse

A total of 12 studies examined anal intercourse. In all, 11 studies examined anal intercourse in heterosexual pornography. One VHS study [37] reported a mean of 1.2 instances in pornographic videos; meanwhile, other VHS studies reported anal intercourse in 10% [35], 13% [41], and 53% [36] of videos. DVD studies found that 42% to 56% [17,44,45] of sex scenes in DVD films included anal intercourse, while another DVD study [46] found a mean of 1.9 acts per film. Internet studies found estimates of 13% to 32% [20,48-50] of videos [20,48-50].

There were 3 studies that examined anal intercourse within gay male pornography. In DVD pornography, anal intercourse was depicted in 80% of sex scenes in one study [45] and 6.8 instances per film in another [46]. Downing et al found that 70% of internet gay male videos depicted anal intercourse [47].

Meanwhile, 4 studies examined ass-to-mouth (ATM). In their study of DVDs, Sun et al [44] found ATM in 44% of male-directed DVD scenes and 41% of female-directed scenes. Specific DVD studies found 22% [45] to 41% [17] of heterosexual scenes depicting ATM, compared with 0% [45] of gay male scenes. One study found ATM in 4% internet video scenes [50].

Group Sexual Activity

Overall, 8 studies examined group sex behaviors. Garcia and Milano [37] found a mean of 1.2 multimale and single female acts, 1.3 multifemale and single male acts, and 0.4 multimale and multifemale acts per VHS film. Brosius et al found 13% of sex scenes featured a man with 2 women, 6% involved 2 men

and 1 woman, and 5% had more than 2 persons of each gender [41]. DVD studies found group sex in 0% to 15% of scenes [17,44]. One study examined gay male pornography, finding 20% of Web-based videos [47] involved 3 or more performers. Brosius et al [41] found a mean of 2.3 actors per VHS sex scene, while other studies found a mean of 1.2 men and 1.2 women actors for general or heterosexual DVDs [45], and 2.5 actors for general or heterosexual internet videos [20]. Gay male studies found a mean of 2.7 men and 0.04 women for DVDs [45], and a mean of 2.4 actors (gender unspecified) for internet videos [47].

Three studies specifically examined a specific group sexual act, double penetration (DP). In their study of DVDs, Sun et al [44] found the highest prevalence of vaginal-anal DP (20% of male-directed scenes, 10% female-directed), while 3% of male-directed scenes and 0% female-directed scenes involved anal DP. In their DVD study, Bridges et al [17] found vaginal-anal DP in 18% of scenes and anal DP in 2% of scenes. Only one internet study [51] examined this act, finding DP in 22% feminist videos, 8% queer feminist videos, 21% "for women" videos, and 17% of mainstream videos.

Bondage and Domination

Bondage, domination, and sadomasochism (BDSM) behaviors were examined by 13 studies. Two studies found BDSM in 10% of gay male internet videos [47], 18% queer feminist scenes [51], and 3% mainstream scenes [51].

Domination was specifically examined by 6 studies. Cowan et al [36] found a general dominance theme occurred in 28% of VHS scenes, and noted that among the scenes characterized by themes of dominance, 78% involved the man in a dominant role and 22% involved women in this role (of these, 37% were women dominating women). Another VHS study found dominance in 14% of sex scenes, with men and women in the dominant role about equally as often (43% depicted dominant men, 45% dominant women, 12% depicted both men and women jointly dominating others) [35]. Another VHS study found 19% of scenes involved dominance, of which 74% had a man in the dominant position while 33% had a woman in this position [43]. A final VHS study found that in 39% of vignettes, a male actor ordered a female actor to perform in a certain way [42]. No studies examined dominance in DVD samples.

In all, 3 studies examined dominance in internet videos, with Klaassen et al [27] finding that 56% of scenes involved a dominant individual, with 39% involving a dominant man and 13% involving a dominant woman. Vannier et al [49] found that in 24% of videos, the male actor had control of the pace and direction of the sexual encounter, while 10% had the woman in control and 66% had shared control. Gorman et al [20] found a male dominance theme in 33% of internet videos (dominance by a woman was not presented in the article). Finally, Fritz and Paul [51] found that male actors instructed the action or behavior of a partner in 25% feminist videos, 19% "for women" videos, and 46% mainstream videos. Comparatively, female actors instructed or directed others in 18% feminist videos, 26% "for women" videos and 33% mainstream videos.

Overall, 5 studies specifically examined submissive behaviors. Palys [35] found that in VHS sex scenes depicting domination, 38% depicted men were in the submissive role, 58% depicted submissive women, and 5% involved men and women jointly dominated by others. Cowan et al [36] found submission in 14% of VHS sex scenes. Another study [49] found that when videos depicted nonegalitarian relations (19% total scenes), 27% involved a man in the submissive position while 78% depicted a woman in this position. An internet study [27] found that 43% of videos involved a submissive female actor, compared with 10% depicting a submissive male actor, while another found that in 47% of videos when at least 2 actors were present, a woman was in a submissive role (results were not provided for prevalence of male submission) [20].

Bondage was examined by 8 studies. Studies of videocassettes indicated bondage was infrequent, present in 2% to 3% [36,40,43] of VHS sex scenes, or 38% of sexually aggressive scenes [35]. Cowan et al [36] also found that only 4 out of 14 bondage scenes depicted a bound man (29%); noting that when men were bound, the tone was typically playful and reciprocal, whereas when women were bound the tone was generally aggressive. In a study of VHS and DVDs, McKee [21] found no instances of bondage, although this was only coded if nonconsensual. Sun et al [44] found that in a sample of DVDs, bondage was seen in 5% of male-directed versus 15% of female-directed scenes, while another DVD study found bondage in 7% of scenes [17]. One internet study [48] found 3% of videos depicted bondage, while another [27] found 0% of internet scenes had men being bound or confined while 1% depicted women in this situation.

Paraphilias

Paraphilias of bestiality, incest, and child pornography were examined, although no studies specifically examined these genres. In all, 3 studies examined bestiality, finding no instances in their samples of video cassettes [38] and DVDs [17,42]. One DVD study reported no instances of pedophilic acts in their sample [17,42]. A total of 4 studies examined incest, prevalent in 0% to 3% of scenes [35,36,39,40].

Orgasms

Several studies provided information about male orgasms or ejaculation, although the focus differed by study. A study of VHS found visible ejaculation in 61% [38] of sex scenes. Vannier et al [49] found that 55% to 58% of Web-based heterosexual videos depicted visible ejaculation, whereas Downing et al [47] found that 47% of Web-based gay male videos depicted any ejaculation.

Some studies specifically identified ejaculation on the face or in the mouth. One DVD study [41] found that in 30% DVD sex scenes, men ejaculated onto a woman's face *or* into her mouth. In all, 5 studies specified ejaculation onto the face of a partner. Studies of general or heterosexual DVDs found estimates ranging from 0% to 5% scenes [17,44], a mean of 4.3 instances per DVD film [46]. Gorman et al [20] found that 45% of Web-based videos depicted ejaculation on the face of a woman by a man. Meanwhile, studies of gay male pornography found a mean of <1% instances of ejaculation on the face of a partner in DVD films [46], or 9% of internet videos [47]. Overall, 3

studies specifically examined ejaculation on or in the mouth of a partner. Studies of general or heterosexual DVDs and videos found this in 48% to 62% of sex scenes [17,44]; meanwhile, Downing et al [47] found this in 8% of gay male internet videos.

A total of 5 studies provided details on female orgasms. Brosius et al [41] noted that less than 1% of women clearly experienced an orgasm in VHS pornography. Prince [38] found that 15% of scenes featured a female orgasm, compared with 61% featuring visible male ejaculation. In video and DVD, a male actor had an orgasm 80% of scenes, compared with 15% for female actors [21]. Similarly, in internet pornography, Klaassen and Peter [27] found that 76% of sex scenes depicted a man having an orgasm, compared with 17% of sex scenes depicting female orgasms. Fritz and Paul [51] found female orgasms in 20% feminist videos, 10% "for women" videos and 15% mainstream videos. Comparatively, they found male orgasms in 53% feminist videos, 63% "for women" videos and 61% mainstream videos.

Condom Use

In all, 6 studies explicitly examined condom use in pornography. For heterosexual pornography, condom use was typically rare. Studies of DVDs found condom use in of 5% to 11% [17,44,45] of sex scenes. In internet samples, 2 studies found condom use in 2% of videos [20,49].

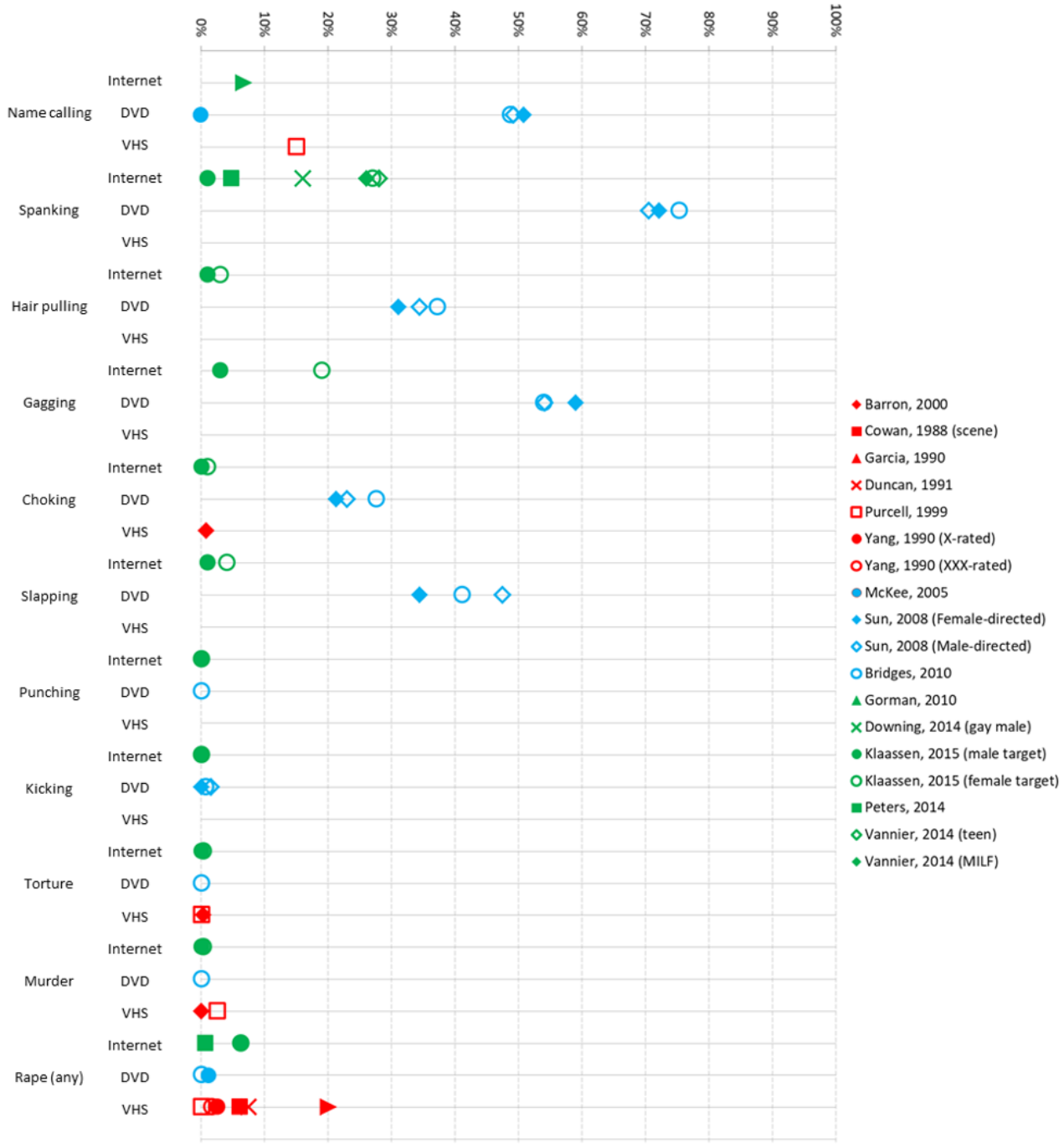
Studies specifically studying gay male pornography had higher rate of condom use than heterosexual pornography. For anal intercourse, condom use was found in 78% of DVD sex scenes [45], while another study found anal intercourse with a condom in 34% of Web-based videos [47]. Only 1 study [47] examined condom use during oral sex between men, finding that nearly 100% was unprotected.

Aggression and Violence

Several studies examined aggression and violence. Of note, only 1 study [35] intentionally included a subsample of violent pornography to form their total sample.

Although many studies present summary variables of both verbal and physical violence, for the purpose of data synthesis, violence is broken down by types of behavior and gender of perpetrator and recipient, when presented. This aims to avoid problems associated with heterogeneity of definitions and behaviors included in these summary statistics. For example, some studies specifically exclude aggressive behaviors which appear to be consensual [21], whereas others purposefully included behaviors regardless of the presence of consent (eg, Garcia and Milano [37]). Most did not specify whether consent was included in their definition. Violent behaviors identified by Monk-Turner and Purcell [42] and Yang and Linz [39] are not presented here as these studies only presented summary statistics. Data are presented graphically in Figure 4.

Figure 4. Percentage of samples depicting key violent behaviors, by medium of pornography (VHS, DVD, or internet-based video). Notes: Results are only presented in this graph if the study presented results by percentage of scenes or videos (ie, studies that presented results by time, mean, character excluded from graph). Only results for specific acts are presented, not for summary variables, due to heterogeneity of these categories. Downing et al's study is presented here as no other gay male/queer studies examined violent behaviors. Palys's (1986) study is not presented here as it purposefully oversampled violent pornography, skewing data. VHS: video home system.



Name Calling and Insulting

A total of 6 studies examined name calling and insulting. Name calling was demonstrated by 6% to 9% of actors [36] and presented toward women in 15% [42] of vignettes in VHS studies. McKee's [21] VHS and DVD study identified no instances of nonconsensual name calling. DVD studies found name calling in 49% to 51% of scenes [17,44]. Name calling was found in 7% of internet videos [20].

Hair Pulling

In all, 4 studies examined hair pulling. In VHS, Cowan and Campbell [36] found that 3% of actors engaged in hair pulling. Two studies examined DVDs; with hair pulling depicted in 31% to 37% of scenes [17,44]. The only internet study [27] found

that 1% of sex scenes involved men having their hair pulled, while 3% involved this for women.

Spanking, Hitting, and Kicking

Spanking was specifically examined by 7 studies. One VHS study found that spanking occurred in 33% of sexually aggressive scenes [35]. McKee's [21] VHS and DVD study identified no instances of nonconsensual spanking. DVD studies found spanking in 71% to 75% scenes [17,44]. Studies of internet pornography had more heterogeneous results. Estimates within general or heterosexual pornography ranged from 5% to 28% [48,49] of videos, while one study found that 27% of sex scenes had women being spanked compared with 1% having men being spanked [27]. Downing et al [47] found spanking in 16% of gay male videos.

In all, 4 studies examined open-handed slapping. McKee's [21] VHS and DVD study identified no instances of nonconsensual slapping. DVD studies found slapping in 34% to 48% of scenes [17,44]. Klaassen and Peter [27] indicated that, in internet videos, 1% of videos involved men being slapped and 4% of videos involved women being slapped.

Punching and kicking were also examined. Overall, 2 studies combined punching and kicking into one variable, with Barron and Kimmel [43] finding this occurred in <1% of VHS scenes, and Palys [35] finding punching or kicking in 24% of aggressive VHS scenes. The 2 studies that specifically examined punching did not find any occurrences in either sample (DVD or internet) [17,27]. In all, 3 studies specifically examined kicking; findings for videos and DVDs ranged from 0% to 2% of scenes [17,44] and, for internet pornography, 0% of videos [27].

Choking and Gagging

A total of 3 studies examined gagging. Of these studies, 2 examined DVDs, finding gagging in 54% to 59% of scenes [17,44]. The remaining study examined internet videos, finding gagging of women occurred in 19% of sex scenes, while gagging of men occurred in 3% of scenes [27].

In all, 4 studies examined choking. Within a VHS sample, Barron and Kimmel [43] found choking occurred in <1% of scenes. Two studies examined DVDs, finding gagging in 21% to 28% of scenes [17,44]. Klaassen and Peter [27] found that, for internet videos, 0% of scenes involved men being choked while 1% involved women being choked.

Torture and Murder

Among 4 studies which examined torture, the oldest study found torture in 2% of aggression scenes [35], and other studies found torture in 0% to 1% of their sample [17,27,42,43]. A total of 5 studies examined attempted or actual murder. For VHS, frequencies ranged from 0% to 3% of scenes [42,43], to 5% [38] of violent acts and 16% of aggression scenes [35]. Other estimates include 0% of DVD scenes [17] and 0% to 1% of internet videos [27].

Rape

The role of consent was examined across several studies. To simplify data synthesis, instances of rape are categorized in accordance with definitions provided by the authors of the studies in the sample. Although not specifically defined by the authors as "rape," estimates by Garcia and Milano [37] and McKee [21] are included in this category due to descriptors including "force" and "nonconsensual" sexual relations. Coercive behaviors are not synthesized due to differences in definitions and conceptual overlap across studies [35-37,39,46].

A total of 10 studies examined rape. Most VHS and DVD studies found rape to be uncommon; depicted in 0% to 8% of scenes [17,21,36,39,40,42] or 0.17 instances per film [37]. However, 1 VHS study, which oversampled violent films, found rape in 31% of sexually violent scenes [35], Cowan et al [36] found that rape occurred in 51% of videos within their sample, and Garcia and Milano [37] found rape in 20% of videos in their sample. One internet study found that "explicit rape" (not defined) occurred in <1% of teen videos [48] while another internet study found rape in 6% scenes [27].

When studies explored rape by gender, most studies found that perpetrators of rape were typically men while victims were typically women. Specifically, Cowan et al [36] found that 90% of VHS instances involving a man raping a woman, while there were 3 instances (<1% scenes) of homosexual rapes (gender unspecified). Another study found that within sexually violent VHS scenes, 29% to 42% involved individual or group heterosexual rape, 3% to 10% involved female homosexual rape and 0% to 3% involved male homosexual rape [39]. Garcia and Milano [37] found 0.2 instances per film of men forcing women into sexual acts, 0.1 instances of female perpetrators and male victims, no instances of male perpetrators with male victims, and 0.1 instances of female perpetrators and female victims. McKee's [21] VHS and DVD study found that among the 1.2% scenes involving rape, there were 5 instances of men forcing women, 2 instances of women forcing men, and one instance of a man and a woman forcing another woman. The only internet study which examined rape by gender found that instances of rape occurred equally for male and female actors [27].

Discussion

Principal Findings

The aim of this study was to systematically review quantitative content analyses of video pornography. The 23 studies reviewed indicate that the pornography in these samples is diverse in terms of behaviors and themes; it is therefore impossible to make definite statements about the content of pornography in general. However, certain trends were noticed when synthesizing these samples.

First, the most common sexual behaviors observed in these are those that some argue to be mainstream, normative or nondeviant in Western culture. For example, in heterosexual pornography, the most common behaviors were fellatio and vaginal intercourse, whereas in in gay male pornography, the most common behaviors were fellatio and anal intercourse. Kissing and cunnilingus were also common (but not universally depicted) in heterosexual pornography, although estimates varied significantly by study. This content is most likely reflecting current sexual norms.

Of note, heterosexual anal intercourse was relatively common in the sample, depicted in 15% to 32% of Web-based videos

and up to half of DVD sex scenes, depending on the sample. A pattern was observed indicating the depiction of heterosexual anal intercourse may have increased since the mid-1980s and the early 1990s, but it has not necessarily increased since the 2000s or with the internet age of pornography. Meta-analytic evidence indicates significant increases in reports of anal intercourse prevalence since the 1970s in some populations [52]. This may be due to lessening of social stigma around anal intercourse [52] but may also be related to the depiction and normalization of anal intercourse in pornography [53]. While heterosexual anal intercourse is enjoyable for many people, for some it is unpleasant and painful; although not studied in this review, it is presumed that pornography does not often show negotiations of consent or preparations for safe anal intercourse.

Paraphilias were very rarely observed in the sample. Bestiality and sexual acts depicting children were not observed in any of the samples. This suggests that these types of acts are very rarely found in general pornography samples, unless specifically searching for this content. While this content is known to be available Web-based, access may require use of the "dark web" or "deep web," parts of the internet not accessible via standard search engines. It is unlikely that people who view easily accessible, free samples (such as those generally included in the content analysis literature) are going to stumble across such content. Regardless, people have reported being exposed to this content; for example, in one study, 18% to 32% of young people reported ever being exposed to child pornography, and 9% to 15% reported ever being exposed to bestiality. However, the number of times participants had seen such images in their lifetime was very low on average, indicating they may have been accidentally exposed [54]. Interestingly, incest was more common than the other paraphilias in the sample, found in three-quarters of the studies which examined it. According to pornography websites, family role play or "faux-cest" videos (note that pornography actors are typically unrelated) increased in popularity between 2014 and 2015, particularly in certain US states [55]. However, PornHub data indicate that "incest" is not among the most popular search terms, but "step mom," "mom," and "step sister" were among the top 20 search terms of 2016 [56]. The implications of this are worth examining in future research.

The results of this study support the public health concern that condoms are rarely depicted in pornography, implying that some actors are at risk of sexually transmitted infection [57]. Condoms were nearly absent in heterosexual samples, and present more frequently in gay or male pornography. Of concern, research has demonstrated that seeing condomless sex in pornography is associated with higher incidence of condomless sex among both populations of heterosexual men [13,58] and men who have sex with men [7]. Although pornography producers have reported concerns that viewers do not wish to see condoms in pornography and are concerned about condoms making sex painful for actors [59], research suggests that most heterosexual men and men who have sex with men are supportive of condom use in pornography [60]. Further research is needed to determine whether such measures would result in meaningful changes in real-life behavior, with considerations of other factors such as perceptions of peers' use of condoms [61].

Contrary to concerns cited by many commentators [19,62], some types of aggression and violence appear to be more common in older forms of pornography compared with the internet pornography. Explicit acts of violence, including rape, appear to be rare in internet mainstream pornography based on the available data. Most studies found that forms of violence such as punching, kicking, torture or murder were only observed a handful of times in the pornography they sampled. However, it is important to note that certain authors (eg, Palys [35]) were able to seek out violent genres of pornography, suggesting that more violent pornography is readily accessible. One study not included in this review demonstrated that it is easy to access internet rape videos if specifically searching for such content [63].

Other forms of aggression appear to be reasonably common, although there was significant variability across estimations. An interesting pattern was observed with spanking, with it not being studied in any VHS studies, being common in 2 DVD studies (ie, Sun et al [44] and Bridges et al [17], who both found spanking in the majority of popular DVDs), while internet studies found spanking in up to a third of content. This pattern suggests that spanking is fairly normalized in mainstream content, although it is difficult to interpret whether spanking is becoming more or less common. Regardless of the nature and frequency of these behaviors, a clear pattern emerged indicating that in general or heterosexual pornography, when aggression and violence occurs, it is more commonly directed toward women, by men. This pattern has also been observed across several studies which used summary measures of violence, which could not be synthesized in this review [17,20,42,44,50].

Other gender inequalities were observed in studies of heterosexual pornography. Orgasm inequalities were universally observed across the literature, with many scenes involving visible male ejaculation and female orgasms rarely depicted (even if they appeared to be fake). Cunnilingus was less common than fellatio, suggesting a pleasure divide. Most dominant actors were men and most submissive actors were women. This suggests that pornography may contribute toward the heterosexual stereotypes that men should dominate or lead sexual activity, that women should be willing to engage in whichever acts are desired by the man, and that both men and women will find such roles normal and enjoyable. Although not able to be systematically reviewed due to differences in definitions and paucity in the literature, studies also noted that unequal power dynamics were fairly common the samples. For example, Monk-Turner and Purcell [42] found that 19% of VHS vignettes involved marked status inequalities based on age, role or occupation in a manner that favors the man. An internet study found that men were more likely to be depicted as having sex for their own pleasure and enjoyment than were women (94 vs 85% of Web-based video scenes) [27]. More studies are needed to explore these dynamics systematically. Further, acts that have been argued to be degrading or demeaning [64] were not uncommon. Ejaculation onto a woman's face or into her mouth was often observed. Ass to mouth penetration was observed in around a third of DVD studies that examined this behavior.

Context of Findings

The varied prevalence of violence and degrading acts in this sample is at least partially influenced by different definitions of violence. The role of consent, which was not systematically studied in this review, is somewhat controversial in the literature. For example, McKee [21] only counted aggressive acts if they appeared to be consensual, arguing that as acts are consensual acts if they do not intend to do another person harm, and the recipient is not motivated to avoid such acts. A distinction is drawn between explicit consent versus coercing another participant into submission [21]. However, Bridges [17] argues that representations of consensual aggressive acts may lead viewers to engage in such acts in a real life, nonconsensual manner. Further, Dines [65] argues that pornography may "hijack" a viewer's sexuality. Despite claims perpetuated by the industry that female actors experience pleasure when participating in a range of sexual acts, the representation of pleasure in pornography can be both complex and ambiguous [22]. Debates about the agency versus exploitation of actors make interpreting these images difficult for audiences and researchers alike [66]. This is important when interpreting findings of content analysis studies to support hypotheses about the impacts of mainstream pornography.

These findings can be considered within the context that pornography viewers are not necessarily passive consumers; opinions of pornography range from extremely negative to extremely positive [4,12,54]. Viewers generally set standards about what they find appropriate or objectionable and likely expose themselves to media that is congruent with their values [67]. For example, it is unlikely that someone opposed to violence against women will purposefully seek out pornography depicting such acts. Regardless, it is possible that some viewers may unintentionally be exposed to and internalize the messages described below, and pornography may subtly shift ideas about which sexual acts are normal, pleasurable, and what their partners expect and desire.

Many viewers are critical consumers of pornography and distinguish between pornography and real-life behaviors. However, other studies have acknowledged that some young people do not demonstrate such critical awareness. For example, one qualitative study [12] of young men identified that many young men have seen pornography depicting violence, even if they did not deliberately seek it out. These viewers may continue watching these videos if they perceive the recipient of the violence is consenting. However, these participants perceived extreme acts of violence and degradation to be normal in this media, speaking about them with detachment and acceptance. The study's authors argued that pornography potentially reinforcing these viewers' understandings of masculine and feminine sexual roles [12].

Limitations

This systematic review only synthesizes a small portion of available pornography. In total, the content analyses in our sample analyzed over 8000 scenes and over 5000 videos. It is difficult to estimate how much pornography is available, although it is estimated that 4% to 15% of internet use involves pornography [68]. The types of pornography analyzed in the studies do not necessarily reflect the types of pornography consumed by viewers. Studies that purposefully sampled more "popular" videos may thereby be more relevant than those that did not, particularly for viewers who purposefully seek out certain videos instead of using random searching to find pornography to watch.

Although most studies in the sample were of moderate to high quality, differences in definitions and methodologies across studies made synthesis challenging. Few details were provided regarding coder characteristics and many studies did not use best practice estimates of interrater reliability [31]. Several studies, particularly older studies, used nonrandom sampling or studied very limited populations. This is understandable as collecting a random sample of VHS pornography would have been very difficult at this time. Sampling methodologies will continue to be a challenge in the future if trends in pornography use and mediums change. Without clear estimates about the number of pornographic videos available, or one source of all videos, it will be impossible to ever obtain a true random sample of internet pornography.

Some studies reported much higher estimates of certain behaviors than others. For example, 2 DVD studies by Bridges et al [17] and Sun et al [44] reported much higher estimations of violent and aggressive behaviors than other DVD studies, or internet-based studies. Further, one study reported a very high prevalence of rape in their sample of video cassettes due to oversampling violent videos [35]. This illustrates how different operationalizations and sampling methodologies can result in very high estimations of certain acts [22].

The authors acknowledge the limitations of this systematic review. It was difficult to identify all pornography content analyses, particularly as some older studies do not use this term. Although the bibliography search attempted to compensate for this, it is possible that the systematic search missed some studies. However, any articles not identified in the search would have been missed systematically [25]. Further, the authors created a custom risk of bias assessment tool due to no appropriate existing tool being identified. Although this tool was based on a pre-existing tool and adapted collaboratively by the authors, it has not been published.

Conclusions

The major implication of this study is that video-based pornography is diverse in its depiction of sexual behaviors and relationships. Owing to there being only being a small number of studies, and inconsistent methodologies, it is difficult to ascertain whether video-based pornography has substantially changed over time. The results tentatively suggest that heterosexual anal intercourse is more common in DVD and internet pornography compared with VHS pornography, several aggressive behaviors (name calling, spanking, hair pulling, gagging, choking, and slapping) are more common in DVD samples than internet and VHS samples, and rape is more common in VHS pornography than newer mediums.

This systematic review indicates that "normative" sexual behaviors are the most frequently depicted in pornography, while extreme violence is rare. Condom use was rare, although

more commonly depicted in gay male pornography. More minor forms of aggression, such as spanking, were more common, and unequal sexual relations (eg, bondage and domination) were also common in general or heterosexual pornography. These behaviors were nearly ubiquitously directed toward women and men were usually depicted as dominant over women. This suggests that gender inequalities are common in pornography, which has implications for the development of healthy sexual relationships among pornography's viewers. Higher quality research, including study replication and consistent methodologies between studies, is needed.

Acknowledgments

The authors would like to thank Michelle Raggatt for her assistance with quality scoring of manuscripts.

Conflicts of Interest

None declared.

References

1. Lim MS, Agius PA, Carrotte ER, Vella AM, Hellard ME. Young Australians' use of pornography and associations with sexual risk behaviours. Aust N Z J Public Health 2017 Aug;41(4):438-443. [doi: 10.1111/1753-6405.12678] [Medline: 28664609]
2. Wright PJ. Pornography and the sexual socialization of children: current knowledge and a theoretical future. J Child Media 2014;8(3):305-312. [doi: 10.1080/17482798.2014.923606]
3. Rothman EF, Kaczmarsky C, Burke N, Jansen E, Baughman A. 'Without Porn … I Wouldn't Know Half the Things I Know Now': a qualitative study of pornography use among a sample of urban, low-income, black and Hispanic youth. J Sex Res 2015;52(7):736-746 [FREE Full text] [doi: 10.1080/00224499.2014.960908] [Medline: 25350847]
4. Davis AC, Temple-Smith MJ, Carrotte E, Hellard ME, Lim MS. A descriptive analysis of young women's pornography use: a tale of exploration and harm. Sex Health 2019 Dec 16:- [Epub ahead of print]. [doi: 10.1071/SH19131] [Medline: 31837713]
5. Harkness EL, Mullan B, Mullan BM, Blaszczynski A. Association between pornography use and sexual risk behaviors in adult consumers: a systematic review. Cyberpsychol Behav Soc Netw 2015 Feb;18(2):59-71. [doi: 10.1089/cyber.2014.0343] [Medline: 25587721]
6. Morgan EM. Associations between young adults' use of sexually explicit materials and their sexual preferences, behaviors, and satisfaction. J Sex Res 2011;48(6):520-530. [doi: 10.1080/00224499.2010.543960] [Medline: 21259151]
7. Schrimshaw EW, Antebi-Gruszka N, Downing MJ. Viewing of internet-based sexually explicit media as a risk factor for condomless anal sex among men who have sex with men in four US cities. PLoS One 2016;11(4):e0154439 [FREE Full text] [doi: 10.1371/journal.pone.0154439] [Medline: 27119990]
8. Marston C, Lewis R. Anal heterosex among young people and implications for health promotion: a qualitative study in the UK. BMJ Open 2014 Jul 18;4(8):e004996 [FREE Full text] [doi: 10.1136/bmjopen-2014-004996] [Medline: 25122073]
9. Vandenbosch L. Antecedents of adolescents' exposure to different types of sexually explicit Internet material: a longitudinal study. Comput Hum Behav 2015;50:439-448. [doi: 10.1016/j.chb.2015.04.032]
10. The Economist. 2015 Sep 26. Naked Capitalism URL: https://www.economist.com/international/2015/09/26/naked-capitalism [accessed 2017-06-14]
11. Walker S, Temple-Smith M, Higgs P, Sanci L. 'It's always just there in your face': young people's views on porn. Sex Health 2015 Jun;12(3):200-206. [doi: 10.1071/SH14225] [Medline: 25935169]
12. Antevska A, Gavey N. 'Out of sight and out of mind': detachment and men's consumption of male sexual dominance and female submission in pornography. Men Masc 2015 Apr;18(5):605-629. [doi: 10.1177/1097184x15574339]
13. Wright PJ, Tokunaga RS, Kraus A. A meta-analysis of pornography consumption and actual acts of sexual aggression in general population studies. J Commun 2016;66(1):183-205. [doi: 10.1111/jcom.12201]
14. Ybarra ML, Mitchell KJ, Hamburger M, Diener-West M, Leaf PJ. X-rated material and perpetration of sexually aggressive behavior among children and adolescents: is there a link? Aggress Behav 2011;37(1):1-18. [doi: 10.1002/ab.20367] [Medline: 21046607]
15. Hald GM, Malamuth NM, Yuen C. Pornography and attitudes supporting violence against women: revisiting the relationship in nonexperimental studies. Aggress Behav 2010;36(1):14-20. [doi: 10.1002/ab.20328] [Medline: 19862768]
16. Riffe D, Freitag A. A content analysis of content analyses: twenty-five years of Journalism Quarterly. Journal Mass Commun Q 1997;74(4):873-882. [doi: 10.1177/107769909707400414]
17. Bridges AJ, Wosnitzer R, Scharrer E, Sun C, Liberman R. Aggression and sexual behavior in best-selling pornography videos: a content analysis update. Violence Against Women 2010 Oct;16(10):1065-1085. [doi: 10.1177/1077801210382866] [Medline: 20980228]
18. Fight the New Drug. 2017 Aug 23. How Consuming Porn Can Lead To Violence URL: https://fightthenewdrug.org/how-consuming-porn-can-lead-to-violence/ [accessed 2017-05-24]

19. McKenzie S. The Sydney Morning Herald. 2011 Mar 11. Why the New 'Porn Norm' is Hurting Women URL: https://www.smh.com.au/politics/federal/why-the-new-porn-norm-is-hurting-women-20110302-1be54.html [accessed 2020-03-05]
20. Gorman S, Monk-Turner E, Fish JN. Free adult internet web sites: how prevalent are degrading acts? Gend Issues 2010;27(3-4):131-145. [doi: 10.1007/s12147-010-9095-7]
21. McKee A. The objectification of women in mainstream pornographic videos in Australia. J Sex Res 2005 Nov;42(4):277-290. [doi: 10.1080/00224490509552283] [Medline: 19827232]
22. McKee A. Methodological issues in defining aggression for content analyses of sexually explicit material. Arch Sex Behav 2015 Jan;44(1):81-87. [doi: 10.1007/s10508-013-0253-3] [Medline: 24609608]
23. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Hoboken, New Jersey: John Wiley & Sons; 2011.
24. 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):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
25. Manganello J, Blake N. A study of quantitative content analysis of health messages in US media from 1985 to 2005. Health Commun 2010 Jul;25(5):387-396. [doi: 10.1080/10410236.2010.483333] [Medline: 20677043]
26. Lim MS, Carrotte ER, Hellard ME. The impact of pornography on gender-based violence, sexual health and well-being: what do we know? J Epidemiol Community Health 2016 Jan;70(1):3-5. [doi: 10.1136/jech-2015-205453] [Medline: 26022057]
27. Klaassen MJ, Peter J. Gender (in)equality in internet pornography: a content analysis of popular pornographic internet videos. J Sex Res 2015;52(7):721-735. [doi: 10.1080/00224499.2014.976781] [Medline: 25420868]
28. The Joanna Briggs Institute. 2017. Checklist for Analytical Cross Sectional Studies URL: https://joannabriggs.org/sites/default/files/2019-05/JBI_Critical_Appraisal-Checklist_for_Analytical_Cross_Sectional_Studies2017_0.pdf [accessed 2017-08-09]
29. Jordan A, Kunkel D, Manganello J, Fishbein M. Media Messages and Public Health: A Decisions Approach to Content Analysis. Abingdon, Oxfordshire: Routledge; 2010.
30. Lacy S, Riffe D. Sampling error and selecting intercoder reliability samples for nominal content categories. Journal Mass Commun Q 1996;73(4):963-973. [doi: 10.1177/107769909607300414]
31. Lombard M, Snyder-Duch J, Bracken CC. Matthew Lombard. 2004. Practical Resources for Assessing and Reporting Intercoder Reliability in Content Analysis Research Projects URL: http://matthewlombard.com/reliability/ [accessed 2017-10-11]
32. Greco T, Zangrillo A, Biondi-Zoccai G, Landoni G. Meta-analysis: pitfalls and hints. Heart Lung Vessel 2013;5(4):219-225 [FREE Full text] [Medline: 24364016]
33. Ryan R. Cochrane Consumers and Communication Review Group. 2013 Jun. Cochrane Consumers and Communication Review Group: Data Synthesis and Analysis URL: http://cccrg.cochrane.org/sites/cccrg.cochrane.org/files/public/uploads/Analysis.pdf [accessed 2020-03-05]
34. Séguin LJ, Rodrigue C, Lavigne J. Consuming ecstasy: representations of male and female orgasm in mainstream pornography. J Sex Res 2018;55(3):348-356. [doi: 10.1080/00224499.2017.1332152] [Medline: 28632461]
35. Palys TS. Testing the common wisdom: the social content of video pornography. Can Psychol 1986 Jan;27(1):22-35. [doi: 10.1037/h0079859]
36. Cowan G, Campbell RR. Racism and sexism in interracial pornography. Psychol Women Q 1994;18(3):323-338. [doi: 10.1111/j.1471-6402.1994.tb00459.x]
37. Garcia LT, Milano L. A content analysis of erotic videos. J Psychol Hum Sex 1991;3(2):95-103. [doi: 10.1300/J056v03n02_07]
38. Prince S. Power and pain: content analysis and the ideology of pornography. J Film Video 1990;42(2):31-41 [FREE Full text]
39. Yang N, Linz D. Movie ratings and the content of adult videos: the sex-violence ratio. J Commun 1990;40(2):28-42. [doi: 10.1111/j.1460-2466.1990.tb02260.x]
40. Duncan DF. Violence and degradation as themes in 'adult' videos. Psychol Rep 1991 Aug;69(1):239-240. [doi: 10.2466/pr0.1991.69.1.239] [Medline: 1961802]
41. Brosius H, Weaver JB, Staab JF. Exploring the social and sexual 'reality' of contemporary pornography. J Sex Res 1993 May;30(2):161-170. [doi: 10.1080/00224499309551697]
42. Monk-Turner E, Purcell HC. Sexual violence in pornography: how prevalent is it? Gend Issues 1999;17(2):58-67. [doi: 10.1007/s12147-999-0015-7]
43. Barron M, Kimmel M. Sexual violence in three pornographic media: toward a sociological explanation. J Sex Res 2000 May;37(2):161-168. [doi: 10.1080/00224490009552033]
44. Sun C, Bridges A, Wosnitzer R, Scharrer E, Liberman R. A comparison of male and female directors in popular pornography: what happens when women are at the helm? Psychol Women Q 2008;32(3):312-325. [doi: 10.1111/j.1471-6402.2008.00439.x]

45. Grudzen CR, Elliott MN, Kerndt PR, Schuster MA, Brook RH, Gelberg L. Condom use and high-risk sexual acts in adult films: a comparison of heterosexual and homosexual films. Am J Public Health 2009 Apr;99(Suppl 1):S152-S156. [doi: 10.2105/AJPH.2007.127035] [Medline: 19218178]
46. Salmon C, Diamond A. Evolutionary perspectives on the content analysis of heterosexual and homosexual pornography. J Soc Evol Cult Psychol 2012 May;6(2):193-202. [doi: 10.1037/h0099215]
47. Downing MJ, Schrimshaw EW, Antebi N, Siegel K. Sexually explicit media on the internet: a content analysis of sexual behaviors, risk, and media characteristics in gay male adult videos. Arch Sex Behav 2014 May;43(4):811-821 [FREE Full text] [doi: 10.1007/s10508-013-0121-1] [Medline: 23733156]
48. Peters EM, Morrison T, McDermott DT, Bishop CJ, Kiss M. Age is in the eye of the beholder: examining the cues employed to construct the illusion of youth in teen pornography. Sex Cult 2014;18(3):527-546. [doi: 10.1007/s12119-013-9210-5]
49. Vannier SA, Currie AB, O'Sullivan LF. Schoolgirls and soccer moms: a content analysis of free 'teen' and 'MILF' online pornography. J Sex Res 2014;51(3):253-264. [doi: 10.1080/00224499.2013.829795] [Medline: 24228745]
50. Zhou Y, Paul B. Lotus blossom or dragon lady: a content analysis of 'Asian women' online pornography. Sex Cult 2016;20(4):1083-1100. [doi: 10.1007/s12119-016-9375-9]
51. Fritz N, Paul B. From orgasms to spanking: a content analysis of the agentic and objectifying sexual scripts in feminist, for women, and mainstream pornography. Sex Roles 2017;77(9-10):639-652. [doi: 10.1007/s11199-017-0759-6]
52. Owen BN, Brock PM, Butler AR, Pickles M, Brisson M, Baggaley RF, et al. Prevalence and frequency of heterosexual anal intercourse among young people: a systematic review and meta-analysis. AIDS Behav 2015 Jul;19(7):1338-1360. [doi: 10.1007/s10461-015-0997-y] [Medline: 25618257]
53. Häggström-Nordin E, Hanson U, Tydén T. Associations between pornography consumption and sexual practices among adolescents in Sweden. Int J STD AIDS 2005 Feb;16(2):102-107. [doi: 10.1258/0956462053057512] [Medline: 15807936]
54. Sabina C, Wolak J, Finkelhor D. The nature and dynamics of internet pornography exposure for youth. Cyberpsychol Behav 2008 Dec;11(6):691-693. [doi: 10.1089/cpb.2007.0179] [Medline: 18771400]
55. May G. VICE. 2015 Feb 25. Why Is Incest Porn So Popular? URL: https://www.vice.com/en_us/article/8gdz8k/why-is-incest-porn-so-popular-332 [accessed 2017-06-14]
56. Pornhub Insights. 2017. Pornhub's 2016 Year in Review URL: https://www.pornhub.com/insights/2016-year-in-review [accessed 2017-08-09]
57. Cohen AC, Tavrow P, McGrath MR. Advocacy coalition for safer sex in the adult film industry: the case of Los Angeles county's Measure B. Health Promot Pract 2018 May;19(3):400-410. [doi: 10.1177/1524839917713942] [Medline: 28656775]
58. Luder MT, Pittet I, Berchtold A, Akré C, Michaud PA, Surís JC. Associations between online pornography and sexual behavior among adolescents: myth or reality? Arch Sex Behav 2011 Oct;40(5):1027-1035. [doi: 10.1007/s10508-010-9714-0] [Medline: 21290259]
59. Shachner J. Unwrapped: how the Los Angeles County safer sex in the Adult Film Industry Act's condom mandate hurts performers & violates the First Amendment. Health Matrix Clevel 2014;24:345-375. [Medline: 25112143]
60. Kraus SW, Rosenberg H. Lights, camera, condoms! Assessing college men's attitudes toward condom use in pornography. J Am Coll Health 2016;64(2):139-146. [doi: 10.1080/07448481.2015.1085054] [Medline: 26629805]
61. Wright PJ, Tokunaga RS, Kraus A. Consumption of pornography, perceived peer norms, and condomless sex. Health Commun 2016 Aug;31(8):954-963. [doi: 10.1080/10410236.2015.1022936] [Medline: 26752322]
62. Long J. The Washington Post. 2016 May 27. Pornography is More Than Just Sexual Fantasy. It's Cultural Violence URL: https://www.washingtonpost.com/news/in-theory/wp/2016/05/27/pornography-is-more-than-just-sexual-fantasy-its-cultural-violence/ [accessed 2020-03-05]
63. Gossett JL, Byrne S. 'Click here': a content analysis of internet rape sites. Gend Soc 2002;16(5):689-709. [doi: 10.1177/089124302236992]
64. Dines G. The white man's burden: Gonzo pornography and the construction of black masculinity. Yale J Law Fem 2006;18:283 [FREE Full text]
65. Dines G. Pornland: How Porn Has Hijacked Our Sexuality. Boston: Beacon Press; 2010.
66. White AG. Pornographic becomings: Female experiences in pornography beyond the victim/agent divide. Bachelor of Arts Honours Thesis, University of Melbourne, Melbourne 2010.
67. Bryant C. Adolescence, Pornography and Harm. Canberra: Australian Institute of Criminology; 2009.
68. Castleman M. Psychology Today. 2016 Nov 3. Dueling Statistics: How Much of the Internet Is Porn? URL: https://www.psychologytoday.com/intl/blog/all-about-sex/201611/dueling-statistics-how-much-the-internet-is-porn [accessed 2017-08-09]

Abbreviations

ATM: ass-to-mouth
BDSM: bondage, dominance, and sadomasochism
DP: double penetration
F-M: female to male
MILF: mother(s) I'd like to fuck

M-M: male to male
VHS: video home system

Edited by G Eysenbach; submitted 15.10.19; peer-reviewed by A Mabry-Flynn, J Zhou; comments to author 08.12.19; revised version received 17.12.19; accepted 26.01.20; published 14.05.20.

Please cite as:
Carrotte ER, Davis AC, Lim MSC
Sexual Behaviors and Violence in Pornography: Systematic Review and Narrative Synthesis of Video Content Analyses
J Med Internet Res 2020;22(5):e16702
URL: https://www.jmir.org/2020/5/e16702
doi:10.2196/16702
PMID:32406863

Review

Supporting Self-Management of Cardiovascular Diseases Through Remote Monitoring Technologies: Metaethnography Review of Frameworks, Models, and Theories Used in Research and Development

Roberto Rafael Cruz-Martínez[1], MSc; Jobke Wentzel[1,2], PhD; Rikke Aune Asbjørnsen[1,3], MSc; Peter Daniel Noort[4], MSc; Johan Magnus van Niekerk[1], MSc; Robbert Sanderman[1,5], PhD; Julia EWC van Gemert-Pijnen[1], PhD

[1]Department of Psychology, Health and Technology, Faculty of Behavioural, Management and Social Sciences, Technical Medical Centre, University of Twente, Enschede, Netherlands

[2]Saxion University of Applied Sciences, Deventer, Netherlands

[3]Research and Innovation Department, Vestfold Hospital Trust, Tønsberg, Norway

[4]Embedded Information Services, Library, ICT Services & Archive, University of Twente, Enschede, Netherlands

[5]GZW-Health Psychology–GZW-General, University Medical Center Groningen, University of Groningen, Groningen, Netherlands

Corresponding Author:
Roberto Rafael Cruz-Martínez, MSc
Department of Psychology, Health and Technology
Faculty of Behavioural, Management and Social Sciences, Technical Medical Centre
University of Twente
Cubicus Bldg
10 De Zul
Enschede,
Netherlands
Phone: 31 068 3186149
Fax: 31 053 4892895
Email: r.cruzmartinez@utwente.nl

Abstract

Background: Electronic health (eHealth) is a rapidly evolving field informed by multiple scientific disciplines. Because of this, the use of different terms and concepts to explain the same phenomena and lack of standardization in reporting interventions often leaves a gap that hinders knowledge accumulation. Interventions focused on self-management support of cardiovascular diseases through the use of remote monitoring technologies are a cross-disciplinary area potentially affected by this gap. A review of the underlying frameworks, models, and theories that have informed projects at this crossroad could advance future research and development efforts.

Objective: This research aimed to identify and compare underlying approaches that have informed interventions focused on self-management support of cardiovascular diseases through the use of remote monitoring technologies. The objective was to achieve an understanding of the distinct approaches by highlighting common or conflicting principles, guidelines, and methods.

Methods: The metaethnography approach was used to review and synthesize researchers' reports on how they applied frameworks, models, and theories in their projects. Literature was systematically searched in 7 databases: Scopus, Web of Science, EMBASE, CINAHL, PsycINFO, Association for Computing Machinery Digital Library, and Cochrane Library. Included studies were thoroughly read and coded to extract data for the synthesis. Studies were mainly related by the key ingredients of the underlying approaches they applied. The key ingredients were finally translated across studies and synthesized into thematic clusters.

Results: Of 1224 initial results, 17 articles were included. The articles described research and development of 10 different projects. Frameworks, models, and theories (n=43) applied by the projects were identified. Key ingredients (n=293) of the included articles were mapped to the following themes of eHealth development: (1) it is a participatory process; (2) it creates new infrastructures for improving health care, health, and well-being; (3) it is intertwined with implementation; (4) it integrates theory, evidence, and participatory approaches for persuasive design; (5) it requires continuous evaluation cycles; (6) it targets behavior change; (7) it targets technology adoption; and (8) it targets health-related outcomes.

Conclusions: The findings of this review support and exemplify the numerous possibilities in the use of frameworks, models, and theories to guide research and development of eHealth. Participatory, user-centered design, and integration with empirical evidence and theoretical modeling were widely identified principles in the literature. On the contrary, less attention has been given to the integration of implementation in the development process and supporting novel eHealth-based health care infrastructures. To better integrate theory and evidence, holistic approaches can combine patient-centered studies with consolidated knowledge from expert-based approaches.

Trial Registration: PROSPERO CRD42018104397; https://tinyurl.com/y8ajyajt

International Registered Report Identifier (IRRID): RR2-10.2196/13334

(J Med Internet Res 2020;22(5):e16157) doi:10.2196/16157

KEYWORDS

eHealth; telemedicine; development; implementation; evaluation; multidisciplinary; qualitative evidence synthesis; meta-ethnography; systematic review; remote monitoring; self-management; cardiovascular diseases; framework; model; theory

Introduction

Holistic Electronic Health Research and Development

Electronic health (eHealth) is the use of technology to support health, well-being, and health care [1]. Research and development approaches in this field are derived from multiple disciplines such as computer, medical, behavioral, and design sciences [2]. Because of this, the use of different terms and concepts to explain the same phenomena [3] and the lack of standardization in reporting interventions [4] often leaves a gap that hinders accumulation of knowledge across the field.

A holistic approach—to recognize the importance of the whole and the interdependence of its parts [5]—to eHealth development could bridge this gap. In practice, this proposes that the interaction and reciprocal influence between contextual, technological, and human factors should be emphasized early and often during eHealth development and be informed by multidisciplinary perspectives [5,6].

Frameworks, Models, and Theories for Development and Design

Researchers and practitioners of eHealth are frequently engaged in iterative phases of development that entail various activities aimed at increasing the success of an intervention (eg, exploring user preferences or assessing usability). These approaches to development are often grounded in a wide and overwhelming variety of frameworks, models, theories, guidelines, or methods [2] that extend across and within scientific disciplines. For example, a recently published framework for development of digital behavior interventions in health care outlines a process of conceptualization, development, implementation, and promotion [7]. In terms of holistic approaches, guidelines have also been used for research and development of eHealth [5,6]. Similarly, some frameworks focus on specific aspects or outcomes such as the adoption, scalability, and sustainability of health care technologies [8]. System development approaches like these are characterized by key ingredients that inform or guide eHealth research, development, implementation, or evaluation. Key ingredients can be principles, guidelines, assumptions, constructs, quality criteria, or grounding themes and ideas.

However, eHealth development is not just about the system's development but also the decision-making process of designing intervention content or features. To this end, system development approaches are often paired with models or theories that, through their own key ingredients, describe knowledge or phenomena or propose predicted relations of constructs. Such theoretical models can be used to understand what mechanism of the intervention works best for whom, instead of evaluating the intervention as a whole (ie, as in traditional effectiveness trials). Key ingredients from theoretical models often relate to behavior change, technology adoption, or the improvement of health [9,10]. While behavior change and improvement of health are typical ingredients of face-to-face interventions, technology adoption is seen as an important outcome of eHealth implementation (because the mode of delivery is often unfamiliar to the user and continued use must be promoted). For example, the persuasive systems design model proposes means for selecting persuasive technology features (ie, promoting adoption) that work best to help users reach their own personal goals (eg, change behavior or improve health) [11]. Considering this type of knowledge, it is plausible that analyzing the use of frameworks, models, or theories applied to system development or intervention design could be a way to understand the potential of eHealth applications to complex health care issues. This focus could enable an understanding of how different approaches can interplay to increase the impact and uptake of health care technology.

Case Study: Supporting Self-Management of Cardiovascular Diseases Through Remote Monitoring Technologies

The integration of self-management support for cardiovascular diseases (CVDs) in technology-based interventions is a case study of interest because it is a contemporary phenomenon, with boundaries not yet clearly evident, in need of answers to questions of how (ie, how does it work?) and why (ie, why does it work?) [12]. In general, CVDs constitute an alarming worldwide health care challenge [13,14] where the provision of remote self-management support is proposed to lessen their burden [15]. Because of the dynamic and continuous (daily) nature of self-management [15,16], monitoring technologies such as internet-enabled blood pressure monitors and weight scales have arguably become a prerequisite of remote care interventions. This approach can also go beyond simply

monitoring by integrating mobile apps or wearables that provide coaching on self-management [17].

The potential of technology-supported remote care relies on real-time monitoring by health care teams, which could facilitate the early detection of symptom deterioration and provision of personalized and timely feedback [15] and also create a feeling of safety for patients who know they are being monitored [18]. Accumulated evidence backs up this potential, as metareviews have shown that technology-supported interventions can be at least as effective as usual care in the promotion of self-management for chronic conditions [19,20].

The complexity of CVD self-management as a health care problem and the required adaptability of the proposed technological solutions [21] naturally cause research in this area to be spread across literature from medical, behavioral, and other disciplines. For example, the expert assessment and recommended treatment of a health care provider (medical science) must be translated into motivational prompts or messages that promote adherence to a desired behavior (behavioral science) and delivered in a salient and understandable way to the patient (human-technology interaction science). Thus, the multidisciplinary gap that characterizes eHealth could also be present in the case of CVD applications.

By and large, it can be recognized that much is generally known about eHealth development processes, based on the many examples that exist, and that research in self-management and CVD is at least equally extensive. Frameworks, models, and theories can be valuable sources of knowledge but are numerous and difficult to compare. Therefore, what is lacking at this point is an analysis that relates the knowledge across different disciplines, the variety of development approaches, and their operationalization in specific settings such as eHealth applications to self-management of CVD.

Aim and Focus

The aim of this review was to identify and compare the underlying approaches that have contributed to the research and development of interventions focused on self-management support of CVDs through the use of remote monitoring technologies. The focus was on frameworks, models, and theories and the reported operationalization of their key ingredients in published studies. The review aimed to create a synthesis that highlighted common (or conflicting) principles, guidelines, and methods to create an understanding of the distinct approaches that have been used for the case at hand.

The review seeks to answer the following research questions. First, what frameworks, models, or theories have been used to develop, implement, or evaluate interventions to support self-management of patients with CVD through the use of remote monitoring technologies? Second, what are the key ingredients of these frameworks, models, or theories that inform or guide the system's (1) development, implementation, or evaluation; (2) content design to promote behavior change and technology adoption; and (3) proposed effects in terms of health-related outcomes? Third, to what extent do the key ingredients of these frameworks, models, or theories fit with the principles of the holistic research and development approach of eHealth [5,6]?

Selecting Metaethnography

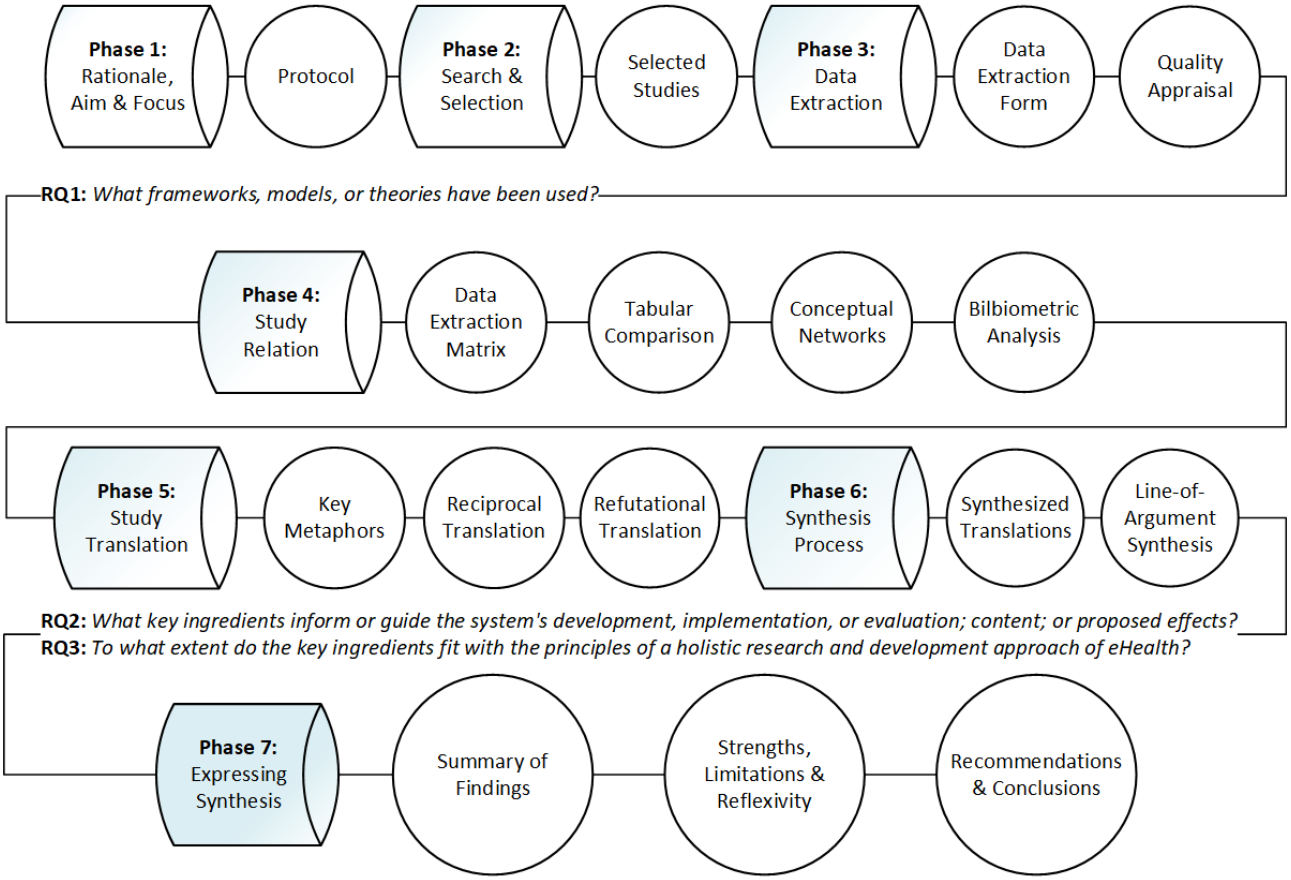
The review was based on metaethnography, a qualitative evidence synthesis method originally developed by Noblit and Hare [22]. Metaethnography is an interpretive approach that seeks to generate a new understanding of a topic. In practice, it involves open coding and constant comparison of key metaphors, which can be phrases, ideas, concepts, perspectives, organizers, or themes identified across the reviewed studies [22]. Metaethnography was selected by following guidelines on choosing qualitative evidence synthesis methods [23]. Mainly, metaethnography was preferred because it aims for an interpretive examination of themes, relationships, and contradictions in the literature, while preserving the social and theoretical contexts of the original findings [24-27]. Additionally, metaethnography fits perfectly with the study aims because it explicitly compares and analyzes texts with the intention to build a holistic interpretation [22]. The aims and methods of the review were described in more detail in a protocol that was preregistered [CRD42018104397] and published [28], complying with best-practice recommendations [24].

Methods

Review Phases

Figure 1 visualizes the phases and key output of this review in relation to the research questions. Phase 1 generated the published protocol [28]. Phases 2 and 3 sought to answer the first research question regarding the identification of frameworks, models, and theories [28]. Phases 4 to 6 operationalized an answer to the second and third research questions through an interpretive characterization of the key ingredients of frameworks, models, and theories [28]. This paper adheres to the recently developed Meta-Ethnography Reporting Guidance (eMERGe) [27].

Figure 1. Phases and key output of the metaethnography review. RQ: research question.



Search Strategy

A comprehensive search was conducted to find published studies (articles) of interest [28]. The search consisted of a systematic database search. Seven databases were used: Scopus, Web of Science, EMBASE, CINAHL, PsycINFO, Association for Computing Machinery Digital Library, and Cochrane Library. The databases were chosen based on their coverage of various fields of science related to eHealth. Multiple key terms were used as part of a highly structured query consisting of four sets aiming for results about frameworks, models, and theories (set 1), eHealth interventions (set 2), self-management (set 3), and cardiovascular diseases (set 4) [28]. The database search was conducted by the main reviewer (RRCM). The search results were uploaded to EndNote X8 (Clarivate Analytics), and duplicates were eliminated.

Study Selection

The selection was performed using the Covidence Web-based software platform (Veritas Health Innovation Ltd). Records were screened by two reviewers, first by title and abstract, and then by full text. The main reviewer conducted the screening throughout all stages. A co-reviewer (RAA) screened at least 15% of the records at each stage and also those tagged as maybe in Covidence by the main reviewer. While 100% dual review would be ideal [29], the approach taken strived for a balance between thoroughness and speed, which is a common strategy for systematic reviews [30] (for full details about study selection see the protocol [28]). Discrepancies between reviewers were discussed and resolved at every stage. Multimedia Appendix 1 presents the selection criteria.

Reading Studies and Extracting Data

The included articles were uploaded to the qualitative software package ATLAS.ti version 8 (ATLAS.ti Scientific Software Development GmbH). Their content was coded according to the elements of a highly detailed data extraction form created to fit the scope of this review (see Multimedia Appendix 1). The data extraction form was based on the Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth (CONSORT-EHEALTH) checklist v.1.6 [31,32]. The form was piloted and iteratively refined throughout all phases. All coded data were also translated to a single form per study. The main reviewer iteratively read, coded, and updated all of the data extraction output based on input from three co-reviewers (RAA, JW, and JGP). The main reviewer also appraised the quality of studies using items from the Critical Appraisal Skills Program's checklists. These checklists were employed because they are a suggested and frequently used tool for metaethnographies [25,26,33-38]. Quality appraisal is not a strict requirement for metaethnography because the richness and relevance of the content is more important [36], but it is considered good practice and was used to get further familiarized with the studies [26].

Determining How Studies Are Related

The relation of studies was performed at three levels with the aim to provide a deep analysis of the data and its context. At the first level, the frameworks, models, theories, and their key

ingredients were compared in tabular form, along with their definitions and occurrence in studies. Additionally, conceptual networks were created in ATLAS.ti using key metaphors as nodes to visualize and explore potential relations (see example in Multimedia Appendix 2).

At the second level, the characteristics of included studies and their overarching projects were compared in tabular form. Moreover, the general context of the underlying approaches was compared. To do this, a matrix was created to visualize the perceived level of clarity and extent of the extracted data [28].

The third level focused on the sources of the underlying approaches identified or cited in included studies (eg, the cited reference of a framework applied in a study). This approach used an explorative bibliometric analysis to assess the multidisciplinary range of such literature and investigate any potential relationships or trends. The references were snowballed and accompanied with co-citation analysis and topic modeling [39-41]. Snowballing can be an alternative to traditional systematic review methods used to identify literature pertaining to a topic of interest by scanning reference lists of studies [42].

Translating Studies

Key metaphors were systematically translated across studies in order to arrive at concepts that embodied more than one study [43]. The translation in a metaethnography is idiomatic (translating the meaning of the text) rather than literal (word for word), and it must take into account the context of the study [22,43]. This stage is characterized by two types of translation: reciprocal and refutational. Reciprocal translations aim to identify or generate metaphors that can better enable holistic accounts of phenomena [22]. On the other hand, refutational translations aim to give explicit attention to incongruities and inconsistencies in the data [43].

To avoid missing valuable insights, the review collected two types of metaphors distinguished by their source. Primary key metaphors were the key ingredients of frameworks, models, or theories operationalized by the authors of a study. Secondary key metaphors were remarkable phrases, concepts, ideas or perspectives by the authors of a study but not apparently derived from a structured underlying approach. Moreover, to assist the translation process it was decided to use the principles of the Center for eHealth Research (CeHRes) roadmap [5,6]. The roadmap is a guideline for holistic eHealth development that is itself based on a review of multiple frameworks and grounded in the integration of persuasive technology design [11], human-centered design, and business modeling [6]. It proposes five principles: (1) eHealth development is a participatory development process; (2) eHealth development creates new infrastructures for improving health care, health, and well-being; (3) eHealth development is intertwined with implementation; (4) eHealth development is coupled with persuasive design; and (5) eHealth development requires continuous evaluation cycles. The roadmap was only used to confront what has been done in the literature with an initial assumption of principles for a holistic approach. This step was operationalized by collectively characterizing both types of metaphors under one or several principles. This process created five clusters, each representing a principle. In the same manner, metaphors were also characterized to clusters of behavior change, technology adoption, or health-related outcomes if they were identified as potential enablers of intervention effectiveness. Clustering the metaphors allowed reviewers to deal with a smaller amount of metaphors at a time so idiomatic translations were performed under each cluster.

Synthesis Process

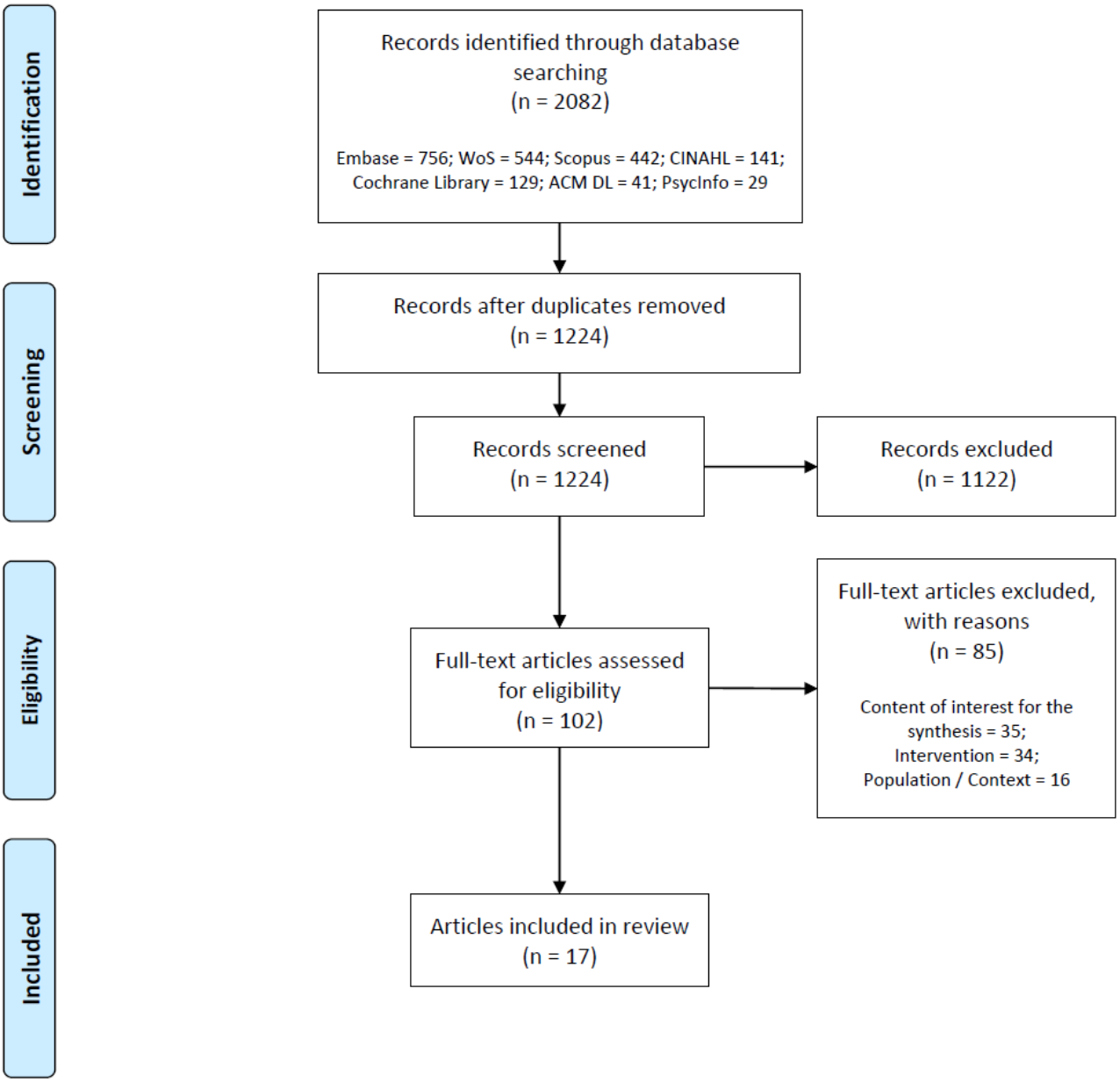
The relation and translation of studies as well as the synthesis process are considered complex analytical and highly overlapping phases of a metaethnography without a one-size-fits-all recipe [43]. To clarify the resulting approach of this review, Multimedia Appendix 2 includes an overview of the activities undertaken in phases 4 to 6 (see Figure 1). From the interpretive findings of the previous phases, a textual line-of-argument synthesis was created. A line-of-argument synthesis is a new storyline or overarching explanation of a phenomenon [43] (the third type of relation after reciprocal and refutational analysis). The synthesis was structured by revising the assumed holistic principles and emphasizing the key metaphors that extended over several themes. Each key metaphor could either contribute to the understanding of specific approaches, highlight important relations across multidisciplinary literature, or even suggest new knowledge derived from integrating unrelated approaches.

Results

Selected Studies

The initial search resulted in 1224 eligible records after removing duplicates. During the title and abstract screening, 1122 records were excluded; 85 more articles were excluded during full-text screening. Of articles excluded by full-text, 41% (35/85) had no relevant content for the synthesis (eg, no apparent framework applied), 40% (34/85) described an intervention that was out of scope (eg, not focused on self-management), and 19% (16/85) had an irrelevant target population or context (eg, not focused on CVD). Multimedia Appendix 3 lists all articles excluded by full-text. In the end, 17 articles were included for the review. Figure 2 summarizes the selection process via the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) flowchart [44]. The keywords, search string, and any additional settings used for the database search are reported as preliminary results in the protocol of this review [28].

Figure 2. Flowchart of included and excluded articles.



Study Characteristics

An overview of the characteristics of included articles can be found in Table 1 (see full table in Multimedia Appendix 4). The articles address 10 different overarching projects, identified as HeartMapp [45-47], Home and Online Management and Evaluation of Blood Pressure (HOME BP) [48-50], Seamless User-Centered Proactive Provision of Risk-Stratified Treatment for Heart Failure (SUPPORT HF) [51-53], PATHway [54,55], Congestive Heart Failure—Personalized Self-Management System (CHF PSMS) [56], MedFit [57], Smartphone Medication Adherence Stops Hypertension (SMASH) [58], Engage [59], MyHeart [60], and a mock-up [61] standalone study. The year of publication of the articles ranged from 2009 to 2018. The United Kingdom (n=7) and United States (n=5) were the most common affiliations of the authors. The most frequent journals in which the articles were published are the Journal of Medical Internet Research (3/17), BMC Medical Informatics and Decision Making (2/17), and Applied Nursing Research (2/17).

The articles were also divided across three target conditions: heart failure (9/17), hypertension (4/17), and CVDs in general (4/17).

Study design classification was done according to the Oxford Centre for Evidence-Based Medicine [62]. Analytic experimental studies are those in which the researcher manipulates the exposure, allocating subjects to the intervention or exposure group. Analytic observational studies are those in which the researcher simply measures the exposure or treatments of the groups without manipulating the exposure or allocation of subjects. Descriptive (qualitative) studies do not try to quantify the relationship but try to give a picture of what is happening in a population.

A total of 35% (6/17) of articles [46,48,49,52,54,60] focused on describing the development process of an intervention and often generally discussed results from more than one study. In these cases, classification by study design was not applicable. For the remaining articles, study design classification revealed

three types of study designs used for eHealth research: analytic observational [45,51,56,59,61], descriptive qualitative [47,50,53,55,57], and analytic experimental [58]. Multimedia Appendix 5 presents the quality appraisal of selected studies.

Table 1. Characteristics of included articles.

Project	Author; country[a]; year	Journal	Target
MyHeart	Villalba et al [60]; Spain; 2009	Conference publication; International Conference on eHealth, Telemedicine, and Social Medicine	HF[b]
SMASH[c]	McGillicuddy et al [58]; US; 2012	Conference publication; Wireless Health	HTN[d]
CHF PSMS[e]	Bartlett et al [56]; UK; 2014	BMC Medical Informatics and Decision Making	HF
SUPPORT HF[f]	Rahimi et al [51]; UK; 2015	European Heart Journal—Quality of Care and Clinical Outcomes	HF
SUPPORT HF	Triantafyllidis et al [52]; UK; 2015	International Journal of Medical Informatics	HF
SUPPORT HF	Chantler et al [53]; UK; 2016	Digital Health	HF
HOME BP[g]	Band et al [48]; UK; 2016	BMJ Open	HTN
HOME BP	Band et al [49]; UK; 2017	Implementation Science	HTN
HOME BP	Bradbury et al [50]; UK; 2017	BMC Medical Informatics and Decision Making	HTN
HeartMapp	Athilingam et al [45]; US; 2016	Applied Nursing Research	HF
HeartMapp	Athilingam et al [46]; US; 2018a	CIN: Computers, Informatics, Nursing	HF
HeartMapp	Athilingam et al [47]; US; 2018b	Applied Nursing Research	HF
Engage	Srinivas et al [59]; US; 2017	International Journal of Human—Computer Interaction	HF
MedFit app	Duff et al [57]; Ireland; 2018	JMIR Formative Research	CVDs[h]
no project	Baek et al [61]; South Korea; 2018	JMIR Cardio	CVDs
PATHway	Walsh et al [54]; Ireland, Belgium, Italy, Greece; 2018a	Translational Behavioral Medicine	CVDs
PATHway	Walsh et al [55]; Ireland, Belgium; 2018b	Journal of Medical Internet Research	CVDs

[a]Countries are included according to the reported affiliations of the authors.

[b]HF: heart failure.

[c]SMASH: Smartphone Medication Adherence Stops Hypertension.

[d]HTN: hypertension.

[e]CHF PSMS: Congestive Heart Failure—Personalized Self-Management System.

[f]SUPPORT HF: Seamless User-Centered Proactive Provision of Risk-Stratified Treatment for Heart Failure.

[g]HOME BP: Home and Online Management and Evaluation of Blood Pressure.

[h]CVD: cardiovascular diseases (in general).

Frameworks, Models, and Theories Applied to Research and Development

In total, 43 frameworks, models, or theories were identified as underlying approaches of the included studies. Textbox 1 and 2 present all of the identified approaches. Multimedia Appendix 6 includes the list of all key ingredients with their definitions per study. Multimedia Appendix 7 and 8 present the full relation between each underlying approach and the operationalized key ingredients by the included studies. In total, 27 different approaches were used to inform the system's development, implementation, or evaluation [52,58,63-92] (Textbox 1 and Multimedia Appendix 7). In contrast, 16 theoretical models were used to inform the system's content [50,93-113] (Textbox 2 and Multimedia Appendix 8).

Multimedia Appendix 7 shows that approaches to system development undertaken by the included studies often promote a participatory, user-centered approach—for example, the development and evaluation process for mHealth [71] or the person-based approach [81,82]. Several types of user-centered models were also identified [61,87,88,90]. Similarly, some frameworks were used to broaden the designer's perspective. For example, the systems engineering initiative for patient safety 2.0 [85] and the patient work lens for consumer-facing health [80] "encouraged the design team to 'think systems' and 'think bigger,' which in this case meant consideration of patients' long-term goals, overall workload, and integration of self-care recommendations into daily life" [59]. Among some focused approaches were, for example, the business-oriented frameworks applied in the HeartMapp project [47].

Textbox 1. Frameworks and models that informed the system's development, implementation, or evaluation.

- 5E usability approach [63]
- Action research [64,65]
- Agile software development [66]
- Behavior change wheel/capability, opportunity, motivation, and behavior model [68,69]
- Business model canvas
- Business research method [70]
- Development and evaluation process for mHealth [71]
- Goal-directed design [72]
- Holistic patient interaction model [73]
- Intervention mapping [74]
- Iterative design model [58]
- Iterative refinement and patient participatory approach [52]
- Iterative software design process [73]
- Iterative software development [75]
- Medical Research Council's guidance for developing and evaluating complex interventions [76-78]
- Multiphase optimization strategy [79]
- Patient work lens for consumer-facing health [80]
- Person-based approach [81,82]
- Practical guide to usability testing [67]
- Realistic evaluation framework [83,84]
- Startup owner's manual [86]
- Systems engineering initiative for patient safety 2.0 [85]
- User-centered design (ad hoc) [87,88]
- User-centered design [89]
- User-centered design of consumer-facing health information technology [90]
- User-centered design (International Organization for Standardization 1999) [91]
- Usability framework [92]

Textbox 2. Theoretical models that informed the system's content.

- Cognitive load theory [93]
- Cognitive theory of multimedia learning [94]
- Common sense model of self-regulation [95]
- Congratulate, ask, reassure, encourage approach [50]
- Control theory framework for personality—social, clinical, and health psychology [96]
- Domestication of technology theory [97,98]
- Information, motivation, behavioral skills model [99]
- Instructional design approach using a pedagogical agent [100]
- Multidimensional framework for patient and family engagement in health and health [113]
- Normalization process theory [101-103]
- Problem-based learning [104]
- Self-determination theory [105,106]
- Social cognitive theory [107-109]
- Social ecological model
- Technology acceptance model [110]
- Unified theory of acceptance and use of technology model [111,112]

Multimedia Appendix 8 shows the wide variety of theoretical models that were used to inform the system's content in the included studies. For instance, social cognitive theory [107-109] was used to outline the behavioral perspective of three different projects. Additionally, some theories were used to understand the process of technology adoption as an outcome, such as the domestication of technology theory [97,98] or the normalization process theory [101-103]. Comparably, technology acceptance was also analyzed through the unified theory of acceptance and use of technology model [111,112] and the technology acceptance model [110].

In general, the integration of multidisciplinary frameworks was frequent in the included studies and their overarching projects. Multimedia Appendix 9 presents how the overarching projects of included studies and their underlying approaches were compared across several levels. Multimedia Appendix 7 and 8 make evident that the HeartMapp [45-47], HOME BP [48-50], SUPPORT HF [51-53], PATHway [54,55], CHF PSMS [56], MedFit [57], SMASH [58], and MyHeart [60] projects were all informed by a combination of approaches from different areas of science. In contrast, the Engage [59] project focused on macroergonomic sociotechnical frameworks while the mock-up [61] study concentrated on a user-centered design research process. However, comparability across projects was influenced by the clarity and extent of the reported data in the selected articles. Multimedia Appendix 10 exemplifies the differences in clarity across studies and projects, while the full comparative analysis can be observed in Multimedia Appendix 9.

The multidisciplinary-based approach was sometimes an explicit goal of researchers. For example, the MedFit study aimed to adopt a "multidisciplinary approach to development [...] drawing on theories from engineering, computer science, and health psychology" [57]. In this line, frameworks were sometimes used to inspire tailored approaches. The most remarkable case was the guidance from the Medical Research Council (MRC) for developing and evaluating complex interventions [76-78], which informed four projects and in general was cited repeatedly in the included studies (see also its relative importance identified in the bibliometric analysis in Multimedia Appendix 9). However, sometimes how a framework informed another one was not completely clear. For example, the iterative software design approach of the MyHeart project [60] was stated to be informed by goal-directed design [72] and user-centered design [91] principles, but this statement was not elaborated in the selected article. Another example is the iterative refinement and patient participatory approach of the SUPPORT HF project [52], which is informed by action research [64,65] and agile software development [66] frameworks. Although such approach is clearly described, its explicit relation to the underpinning frameworks is not explicitly established.

Key Ingredients That Inform or Guide Development, Content, or Outcomes

The key ingredients as presented in Multimedia Appendix 7 and 8 facilitate a more detailed comparison of how underlying approaches were used. Mainly, the approaches to system development contain key ingredients that mostly represent stages of development, implementation, or evaluation. The stage-based ingredients that focus on creating a fit between the user and the proposed solution (eg, through co-design and formative evaluation) are eHealth-specific frameworks [71,81,82], guidelines for (software) iterative evaluation [52,58,73], or user-centered design methods [87-91]. On the other hand, some stage-based key ingredients guided systematic exploration, selection, and integration of theory with empirical evidence (eg, establishing why or how the intervention works through theoretical modeling). These ingredients are instead derived from research and intervention-building frameworks from behavioral [74], medical [76-78], or sociological [83,84] areas

of science. Other key ingredients did not represent stages of development but were constructs used to broaden the designers' perspectives—for instance, to understand human-technology interaction [73], the patient's work system [80,85] (ie, the workflow), ecosystem levels in health care [55], or key insights for business modeling [47]. The bibliometric analysis on the cited references of underlying approaches also observed a distinction between topics of intervention development, behavior change, and health care (see Multimedia Appendix 9).

In contrast, theoretical models provide key ingredients that were used to inform the content or outcomes of interventions. These ingredients could be psychological determinants [68,69,95,99,105-109] (eg, self-efficacy), mechanisms of action (self-monitoring [96]), or mediators (engagement [113]) for behavior change. Other key ingredients are about eHealth adoption, such as determinants of technology acceptance [110-112] (eg, ease of use) or processes and mediators of adoption [97,98,101-103] (eg, objectification).

In sum, the included studies highlighted participatory, user-centered, and iterative approaches with multiple perspectives about how to effectively influence the uptake of eHealth at several levels (eg, from individual cognition to the elements of a macroergonomic work system). Multimedia Appendix 6 shows how these ingredients and other insights (metaphors) of the included studies were compared and translated within and across studies. In the sections below, the included studies [45-61] (Multimedia Appendix 4) are mentioned by the first author's name in the text, their underlying approaches (Textbox 1 and 2 and Multimedia Appendix 7 and 8) are named and referenced when applicable.

Behavior Change

The effectiveness of eHealth systems in the included studies in terms of behavior change was operationalized by their success in improving self-management behaviors. In this regard, the operationalization of key ingredients could be better understood through the sociotechnical perspective which broadly conceptualizes self-management as a complex biopsychosocial process, as proposed by the systems engineering initiative for patient safety 2.0 [85] and the patient work lens for consumer–facing health [80] model. Throughout the included studies, the proposed general solution was the provision of tailored, personalized, or timely support (Band et al [48]), grounded in the potential of eHealth to deliver behavior change techniques that can facilitate long-term sustained behavior change (Duff et al [57]). Key ingredients were mostly informed by psychological theories such as social cognitive theory, which highlights determinants like self-efficacy, outcome expectancy, individual goals, and perceived impediments and facilitators [107-109]. Likewise, information, motivation, behavioral skills, and social opportunity were also parameters used by the selected studies to facilitate behavior change, based on the behavior change wheel [68,69] or the information, motivation, behavioral skills model [99]. Behavior change was also proposed to be at play during the adoption of a technology according to the normalization process theory [101-103]—for example, to explain how patients or health care providers must integrate several behaviors into everyday life (interactional workability) or how patients must be able to adapt their self-care routines when required (reconfiguration). Finally, the review collected a long list of practical applications (translations of behavior change techniques into intervention components) that showcased the similarity of current approaches to support self-management through remote monitoring technologies. For example, a familiarization phase (Walsh et al [54]) with the technology was an approach used by several studies. The most common features of the technologies included assessment, self-monitoring, feedback (during activity and after performance), behavioral change support (eg, goal setting, promoting home exercising), and education (eg, on disease management).

Technology Adoption

The effectiveness of eHealth systems in the included studies in terms of promoting technology adoption during the implementation process was operationalized mainly by the aim to create a fit between the system and the self-management routines of the patients. Primarily, technology adoption was informed in the included studies by the recognition of a diversity of user experiences (Chantler et al [53]), and the predominant strategy to undertake user-centered design [87,89] to address this heterogeneity. Once again, the tailored, personalized, and timely support (Band et al [48]) was the main driver during operationalization. Specifically, the adaptation to the personal routines of patients (Villalba et al [60]) was identified as a common idea across the included literature. In addition, the inclusion of a bidirectional service model (as in Baek et al [61]) or blended care which entailed communication between health care providers and patients was also an important theme across the included studies. This was in part because the sense of connection to a support team that a system provides to a patient could act as a key motivator for the use of the technology (Chantler et al [53]). Guidelines for health care providers to offer patient-centered support within a remote care context were applied by one of the included studies (Bradbury et al [50]). Remarkably, two major challenges of technology adoption were also identified. First, the technology knowledge gap (literacy) between younger and older generations (discussed by Duff et al [57]). Second, the inertia of disengagement, which was proposed to be tackled by the establishment of design goals that promote rather than assume baseline levels of engagement (Srinivas et al [59]). Technology adoption could also be assessed at multiple levels—for example, through a user interaction model (applied by Villalba et al [60]) that investigates the explicit and implicit interaction between the user and the technology or in terms of a multidimensional usability framework (applied by Walsh et al [54]). Notably, technology adoption could be explored through models such as the domestication of technology theory [97,98], which describes the processes of acceptance, rejection, and use of technology by its users (applied in Chantler et al [53]). Likewise, the unified theory of acceptance and use of technology and the technology acceptance model were other models of adoption that offered determinants such as behavioral intention, performance and effort expectancy, experience, and price value [110-112]. Finally, the key insights for building a minimum viable product (eg, value propositions), derived from the business model

canvas, were also interpreted as key ingredients to enable the desired adoption during implementation of the technology (Athilingam et al [47]).

Health-Related Outcomes

The paths to health improvement of the eHealth systems in the included studies were several. Overall, most of the listed ingredients could be categorized as engagement outcomes (eg, continued use and high usability), behavioral outcomes (eg, improved self-management), or health-related outcomes (eg, reducing admissions or increasing quality of life). In these terms, the operationalization of health-related outcomes in the selected studies focused notably on behavior change as the indirect path to increase health, an approach often grounded in the behavior change wheel and its capability, opportunity, motivation and behavior model [68,69]. For example, technologies were designed to include several intervention functions, such as enablement (increasing means and reducing barriers to perform the behavior), education (increasing knowledge or understanding), and environmental restructuring (changing the physical or social context). Moreover, the sociotechnical perspective of the systems engineering initiative for patient safety 2.0 [85] was used by Srinivas et al [59] to broaden the understanding of the various components of an intervention (eg, work processes) in relation to their impact on potential outcomes (proximal or distal, desirable or undesirable). An important challenge to improve health in the selected literature focused on hypertension (see Table 1 and Multimedia Appendix 4) was clinical inertia [58] (ie, the failure to establish appropriate targets and escalate treatment to achieve treatment goals). Additionally, the accurate measurement of changes in key determinants (eg, knowledge, as approached by Bartlett et al [56]) was also a possible methodological obstacle.

Fit of Key Ingredients With Holistic Principles for Research and Development

Projects at the intersection of self-management, CVD, and eHealth have directly or indirectly applied holistic principles for research and development. Namely, the principle of eHealth as a participatory development process and the principle that eHealth development is intertwined with implementation have been explicitly endorsed in the included studies. On the other hand, the principle that eHealth development creates new infrastructures for improving health care, health, and well-being has been partially operationalized through the use of various frameworks but has remained unacknowledged as a key underlying theme. Similarly, the principle that eHealth requires continuous evaluation cycles has been indirectly addressed by a wide variety of aims and methods operationalized across many phases of the eHealth development process. Ultimately, the principle that eHealth development is coupled with persuasive design was unacknowledged across included studies, although varied and generic approaches to inform design were found. Multimedia Appendix 11 illustrates the terms and definitions (metaphors) that constituted each cluster through conceptual networks.

Development Is a Participatory Process

The principle of participatory development has been widely operationalized as part of a fundamental integration of person-based approaches with theory and evidence (Band et al [49]) and directly grounded in the concept of user involvement, which was promoted throughout the included literature to conform with the guidelines of the MRC [76-78]. Moreover, the participatory approach was complemented with a socioecological perspective to secure inclusion of a diversity of user experiences (Chantler et al [53]) and multiple levels of the target group's ecosystem (as applied by Walsh et al [55]). The aims and methods for participatory development of the included studies have been extensively underpinned by user-centered design [87,89] and applied to the full extent of eHealth development phases (from planning to deployment [80-82,88]).

Development Creates New Infrastructures for Improving Health Care, Health, and Well-Being

Initially, the principle that eHealth development creates new infrastructures for improving health care, health, and well-being was thought to be self-evident given the scope of the review (remote care). The established aims of researchers and developers in the selected studies endorsed this principle, such as providing tailored, personalized, and timely support (Band et al [48]) or the unobtrusive remote delivery of system refinements (Triantafyllidis et al [52]). Key contextual factors were also highlighted by the included studies, such as the facilitating conditions (perceptions of the resources and support available to perform a behavior) defined by the unified theory of acceptance and use of technology model [111,112]. In this regard, as posed by the behavior change wheel [68,69], context can also include the policy categories surrounding technology-supported interventions (decisions made by authorities that help to support and enact an intervention). An early step to create an infrastructure can be to develop a program plan to describe the scope and sequence of intervention components, its required materials, and the protocols for implementation (as in intervention mapping [74]). In addition, the use of interdisciplinary methods (eg, factorial or fractionated evaluation designs discussed by Walsh et al [54]) and a socioecological perspective (Walsh et al [55]) are approaches that can facilitate the understanding of eHealth infrastructures and ecosystems (ie, identifying what works, who should be involved, and how in remote care support).

Development Is Intertwined With Implementation

An implementation focus such as the one promoted by the development and evaluation process for mHealth [71] was prominent across the selected literature, directly supporting this principle. However, the aims and methods to accomplish this were often vaguely and partially described. For example, business modeling [70,86] approaches have been used for research (Athilingam et al [47]), but only for an initial conceptualization of the technology (a first concept of the solution that still requires validation, as defined by the iterative software design process [73]). A highlighted example of development intertwined with implementation was the aim to provide remote delivery of system refinements as proposed in

the iterative refinement and patient participatory approach applied by Triantafyllidis et al [52]. This approach intended to facilitate continuous system updates without the use of valuable human resources. For this principle, only formative research (eg, focus groups in Duff et al [57]) and field studies (eg, Bartlett et al [56]) have been employed as methods that can be intertwined with the development process to understand and ideally increase the uptake of the technology.

Development Integrates Theory, Evidence, and Participatory Approaches for Persuasive Design

The term of persuasive design, prominent in the field of human-computer interaction, was completely omitted in the included literature. However, it was evident that the integration of theory-, evidence- and person-based approaches (Band et al [49]) was used to increase persuasiveness. In other words, the selected studies implicitly set persuasiveness as part of their development aims—for example, by the proposed personalization and tailoring (Chantler al [53]) of the intervention, the creation of habits in the use of a technology, or the leverage on the hedonic (fun, pleasure) (Duff et al [57]) and automatic motivation (emotional reactions) of end users (Band et al [49] or Walsh et al [54]). In this regard, theoretical approaches were often related to theoretical modeling (eg, the logic model of Band et al [49]), while evidence was explored through preclinical or theoretical research (eg, literature reviews) conforming to the MRC's guidelines [76-78]. As mentioned before, participatory or person-based approaches were more often applied as part of user-centered design. Hence, this revised principle highlighted how the included studies coped with the challenge of knowledge translation across different areas of research and its application to a specific aim (ie, integrating multidimensional ingredients that contribute to a common goal). To exemplify this, the aim for personalization and tailoring was derived from evidence that prioritized "the need to tailor...systems to user's capacity and preferences" (Chantler et al [53]), rather than preassuming these aspects as key principles to increase technology persuasiveness via the facilitation of task support (as proposed by the persuasive systems design model [11]). One trade-off made apparent by this revised principle and example is that the identified approaches were not related to theory developed specifically for technology-based interventions, and therefore their application to this area seemed to be open to the interpretation of researchers and developers.

Development Requires Continuous Evaluation Cycles

The requirement of continuous evaluation cycles in eHealth development revealed a contradiction within the included literature. The contradiction was outlined by the MRC's [76-78] proposed stepwise development of complex interventions, as opposite to its own practical recommendation to undertake a parallel approach that can integrate stages with distinct aims into larger phases of development. For example, a large phase of development can include preclinical or theoretical research (eg, understanding the users and their environment through literature reviews) [76-78], early solution finding (eg, discussing solutions with the target group as defined by the iterative design model [58]), and an initial theoretical conceptualization (as defined by the development and evaluation process for mHealth [71]) or modeling of the eHealth technology and its components (eg, deciding on the theoretical basis and proposing how an intervention could work) [76-78]. In practical terms, evaluation cycles were often defined by either the choice of an agile (rapid and cyclical stages) or waterfall approach (long and sequential stages) to product development (as discussed by Srinivas et al [59]). This principle also highlighted the importance of integrating interdisciplinary methods (as proposed by Walsh et al [54]) that accommodate to the planned evaluation cycles. In this regard, the creation of an evaluation plan (as in intervention mapping [74]), where variables are defined in a measurable way in relation with the intervention objectives, methods, and strategies, seemed to be a key phase to bridge early design with formative evaluation processes of eHealth. To apply continuous evaluation cycles, the included studies made wide use of user-centered design methods [87,89] such as usability testing but also other frameworks such as realistic evaluation [84], which is a theory-driven approach to evaluate the complexity of social programs (applied by Bartlett et al [56]).

Discussion

Principal Findings

The findings of this review confirm and exemplify the remarkable challenges posed by the multidisciplinary gap in the field of eHealth. Mainly, the review listed 43 multidisciplinary frameworks, models, theories, and guidelines that have informed interventions within the scope of eHealth applications to self-management of CVD. Multidisciplinary approaches were often integrated and aimed to create a fit between users, the content of an intervention, and its context. The following sections summarize and assess the contributions of the principal findings with prior and related works.

Bridging the Multidisciplinary Gap in Electronic Health Research and Development

In terms of development, the findings of this review place the integration of theory-, evidence- and participatory approaches to inform persuasive design as a newly generated overarching principle [49]. To do this, the studies often integrated knowledge from several disciplines, which in general has been argued as positive and desirable for eHealth [2]. However, in terms of design, one downside from the selected studies was that the approaches considered were often constrained to behavioral or sociological perspectives that were not focused on increasing the use and uptake of technology. In terms of implementation, this review suggests the importance of interdisciplinary methods that integrate broad perspectives such as the socioecological, sociotechnical [80,85], or business modeling [70,86] approaches. Specifically, the importance of workflow for the success of eHealth interventions has also been observed in another review [114]. Workflow can be defined as the way people interact with their work, communication pathways, and other people [114]. The inclusion of novel technological tools in the workflow of patients and health care providers was addressed in the reviewed studies through the lens of models such as the systems engineering initiative for patient safety 2.0 [85] framework or the domestication of technology theory [97,98]. For evaluation of eHealth, the reviewed literature acknowledged the iterative

nature of this process, but some of the identified approaches seemed to still be restrained by fixed stages of postdevelopment testing of effectiveness. It must be noted that these fixed research programs can hinder the adaptability of interventions to the dynamic and flexible reality of the patients [115]. In this light, a previous review on the adoption of self-management solutions has also showed that a broad "consideration of preconceived barriers and facilitators for adoption" might be too simplistic, because what is perceived as a barrier or facilitator for one individual could have the opposite effect for another [116]. To maximize adoption, it is therefore recommended to iteratively reevaluate key social, motivational, cultural, moral, and financial factors [116]. The continuous evaluation of these factors can be matched with participatory and user-centered principles.

Challenge of Reporting Intervention Content and Design

Overall, the findings of this review are in line with the general literature addressing several advantages to the use of theoretical frameworks for eHealth development and design and the different ways in which they can be operationalized [3]. However, the major challenge of adequate reporting of intervention design and content was also recognized (as discussed by Walsh et al [54] and Srinivas et al [59]). The lack of specification of the underlying approaches and their operationalization is still "suspected to be an artifact of publishing conventions and space constraints, as much as if not more than the nature of actual research being performed" [59]. All in all, the review included exemplary cases of publications with rich conceptual and descriptive data about eHealth development and design [46,49,54,56,57,59,60].

Strengths and Limitations

This is considered to be the first metaethnography focused on bridging knowledge from multidisciplinary fields of science to better understand and improve eHealth research and development approaches. The review made great efforts to follow a thorough, systematic, multilevel approach [28], adhere to recently developed guidelines [27], and be informed by similar studies [22,25,26,33-35,37,38,43,117-121]. Although the number of papers included was relatively low, metaethnography is a complex methodology and synthesis process that entails numerous challenges and limitations [122]. For this review, a main limitation in the search phase was that no efforts were made to contact the authors to request additional information on their studies. This would have added additional time constraints that were not seen as feasible. For the same reason, although reference tracking was originally planned [28], no further inclusions through this method were considered. Although it was an exclusion criterion, it could arguably be seen as a limitation that some studies were excluded because they did not explicitly describe their underlying approaches. Including more papers could have arguably added new perspectives to the synthesis, but the added time to the interpretive task would have been too burdensome. In the translation phase, several concepts and themes required a high level of interpretation and study contextualization acquired by rereading the articles several times and with different intentions (eg, for data extraction, comparison, or verification) [25]. The main reviewer applied this approach, but co-reviewers followed a sequential approach focused on validation or identification of inconsistencies. Finally, it should also be noted that the key ingredients were sometimes extracted from the sources cited by the selected studies. Hence, the review could have missed updates and refined assumptions or principles. For example, the intervention mapping protocol has been continuously upgraded in comparison with the cited source of the selected studies [123].

Conclusions and Recommendations

The multidisciplinary gap naturally constrains eHealth research and development to the structures and perspectives of discipline-specific frameworks that often miss key factors of the complex reality in health care. A holistic approach to the problem should consider multidisciplinary principles, such as those outlined by this review, to better define, structure, and report the underlying approaches to research and development of future eHealth interventions. The principles of the CeHRes roadmap mapped fairly well to what has been done in the selected literature. Positively, the use of participatory, user-centered design, and continuous evaluation cycles were commonly applied principles. On the contrary, less attention was given to the integration of implementation in the development process and implications of the new eHealth-based health care infrastructures as a whole. The integration of theory and evidence to inform (persuasive) design was an important principle that arose from the included studies, but the frameworks or models used to this purpose are not focused on creating a fit between human and technology.

Overall, it is recommended that researchers and developers make explicit and concrete statements about their approaches to eHealth. For instance, once a thoughtful decision has been made on guiding frameworks, models, or theories, it would be useful to also underline the holistic principles that are considered valuable by the research team (eg, will the approach consider existing evidence and theory or will it be solely guided by new data?). Unfortunately, there are no gold standards to report the content of eHealth interventions, beyond the CONSORT-EHEALTH checklist v.1.6 [31,32], which is specific to trials, and even less so to report underlying guiding principles. In the future, clearer operationalization (and reporting) of guiding frameworks and theoretical models is seen as vital to advance such knowledge in the field, as better predictive theories could provide an answer to the question "what works, for whom, in what settings, to change what behaviors, and how?" [124]. By and large, both theory and evidence must converge to determine the most effective mechanisms for technology-supported interventions. To accomplish this and move beyond what can be learned from published literature, holistic approaches can integrate patient-centered studies with consolidated knowledge from expert-based approaches (eg, via Delphi or other group decision-making methods [125]).

Finally, many questions still remain regarding the optimal use and advantages of specific frameworks or theories for eHealth development. Future reviews could aim to compare the effectiveness of theory-based eHealth interventions with those that do not make use of any. Moreover, more exploratory work

is needed to understand how different frameworks or theories are more relevant or useful for specific settings and contexts (eg, which types of theories or frameworks are better suited to inform remote care interventions and why?).

Acknowledgments

This review is part of the PhD project of RRCM, funded by the Mexican National Council for Science and Technology (CONACYT, in Spanish).

Authors' Contributions

RRCM led and was involved throughout all stages of the review process and managed the preparation of the manuscript. PDN assisted on the selection of the review method and provided methodological input at various stages. RAA contributed as co-reviewer for the selection process. RAA, JW, and JGP contributed to the translation and synthesis stages (see Multimedia Appendix 2). RS contributed in the ideation and early stages of the review process. JMN conducted the explorative bibliometric analysis. All authors critically evaluated the article and gave their final approval before submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Selection criteria and data extraction form.
[DOCX File , 49 KB - jmir_v22i5e16157_app1.docx]

Multimedia Appendix 2
Overview of relation, translation, and the synthesis process.
[DOCX File , 259 KB - jmir_v22i5e16157_app2.docx]

Multimedia Appendix 3
Excluded articles at full-text screening.
[DOCX File , 35 KB - jmir_v22i5e16157_app3.docx]

Multimedia Appendix 4
Characteristics of included articles.
[DOCX File , 16 KB - jmir_v22i5e16157_app4.docx]

Multimedia Appendix 5
Quality appraisal of included studies.
[DOCX File , 25 KB - jmir_v22i5e16157_app5.docx]

Multimedia Appendix 6
List of identified and translated metaphors.
[DOCX File , 81 KB - jmir_v22i5e16157_app6.docx]

Multimedia Appendix 7
Frameworks and models that informed the system’s development, implementation, or evaluation and their operationalized key ingredients by included studies.
[DOCX File , 19 KB - jmir_v22i5e16157_app7.docx]

Multimedia Appendix 8
Theoretical models that informed the system’s content and their operationalized key ingredients by included studies.
[DOCX File , 15 KB - jmir_v22i5e16157_app8.docx]

Multimedia Appendix 9
Comparison of included projects and their underlying approaches.
[DOCX File , 346 KB - jmir_v22i5e16157_app9.docx]

Multimedia Appendix 10
Perceived level of clarity and extent of the reported data related to development and design in included articles and projects.
[DOCX File , 17 KB - jmir_v22i5e16157_app10.docx]

Multimedia Appendix 11
Conceptual networks of clustered metaphors.
[PDF File (Adobe PDF File), 1077 KB - jmir_v22i5e16157_app11.pdf]

References

1. van Gemert-Pijnen LJ, Kip H, Kelders S, Sanderman R. Introducing eHealth. In: eHealth Research, Theory and Development. London: Routledge; 2018:23-46.
2. Pagliari C. Design and evaluation in eHealth: challenges and implications for an interdisciplinary field. J Med Internet Res 2007;9(2):e15 [FREE Full text] [doi: 10.2196/jmir.9.2.e15] [Medline: 17537718]
3. Hekler E, Klasnja P, Froehlich J, Buman M. Mind the theoretical gap: interpreting, using, and developing behavioral theory in HCI research. CHI13 Proc SIGCHI Conf Hum Factors Comput Syst 2013 Apr:3307-3316. [doi: 10.1145/2470654.2466452]
4. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ 2014 Mar 07;348:g1687. [doi: 10.1136/bmj.g1687] [Medline: 24609605]
5. Kip H, van Gemert-Pijnen LJ. Holistic development of eHealth technology. In: eHealth Research, Theory and Development. London: Routledge; 2018:151-186.
6. van Gemert-Pijnen JEW, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
7. West R, Michie S. A Guide to Development and Evaluation of Digital Behaviour Interventions in Healthcare. London: Silverback Publishing; 2016.
8. 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]
9. Naslund JA, Aschbrenner KA, Kim SJ, McHugo GJ, Unützer J, Bartels SJ, et al. Health behavior models for informing digital technology interventions for individuals with mental illness. Psychiatr Rehabil J 2017 Sep;40(3):325-335. [doi: 10.1037/prj0000246] [Medline: 28182469]
10. Winter SJ, Sheats JL, King AC. The use of behavior change techniques and theory in technologies for cardiovascular disease prevention and treatment in adults: a comprehensive review. Prog Cardiovasc Dis 2016 May;58(6):605-612 [FREE Full text] [doi: 10.1016/j.pcad.2016.02.005] [Medline: 26902519]
11. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Comm Assoc Inf Syst 2009;24(1):28. [doi: 10.17705/1cais.02428]
12. Yin R. Applications of Case Study Research. Thousand Oaks: Sage Publishing; 2011.
13. Roth G, Johnson C, Abajobir A, Abd-Allah F, Abera S, Abyu G, et al. Global, regional, and national burden of cardiovascular diseases for 10 Causes, 1990 to 2015. J Am Coll Cardiol 2017 Jul 04;70(1):1-25 [FREE Full text] [doi: 10.1016/j.jacc.2017.04.052] [Medline: 28527533]
14. World Health Organization. Global status report on noncommunicable diseases 2014: attaining the nine global noncommunicable diseases targets; a shared responsability URL: http://apps.who.int/iris/bitstream/10665/148114/1/9789241564854_eng.pdf?ua=1 [accessed 2020-04-13]
15. Riegel B, Moser DK, Buck HG, Dickson VV, Dunbar SB, Lee CS, American Heart Association Council on CardiovascularStroke Nursing; Council on Peripheral Vascular Disease;Council on Quality of CareOutcomes Research. Self-care for the prevention and management of cardiovascular disease and stroke: a scientific statement for healthcare professionals from the American Heart Association. J Am Heart Assoc 2017 Aug 31;6(9):1 [FREE Full text] [doi: 10.1161/JAHA.117.006997] [Medline: 28860232]
16. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 2002;48(2):177-187. [Medline: 12401421]
17. Kim BY, Lee J. Smart devices for older adults managing chronic disease: a scoping review. JMIR Mhealth Uhealth 2017 May 23;5(5):e69. [doi: 10.2196/mhealth.7141] [Medline: 28536089]
18. Middlemass JB, Vos J, Siriwardena AN. Perceptions on use of home telemonitoring in patients with long term conditions—concordance with the Health Information Technology Acceptance Model: a qualitative collective case study. BMC Med Inform Decis Mak 2017 Jun 26;17(1):89 [FREE Full text] [doi: 10.1186/s12911-017-0486-5] [Medline: 28651588]
19. Greenwood DA, Gee PM, Fatkin KJ, Peeples M. A systematic review of reviews evaluating technology-enabled diabetes self-management education and support. J Diabetes Sci Technol 2017 May 31;11(5):1015-1027. [doi: 10.1177/1932296817713506] [Medline: 28560898]

20. Hanlon P, Daines L, Campbell C, McKinstry B, Weller D, Pinnock H. Telehealth interventions to support self-management of long-term conditions: a systematic metareview of diabetes, heart failure, asthma, chronic obstructive pulmonary disease, and cancer. J Med Internet Res 2017 May 17;19(5):e172 [FREE Full text] [doi: 10.2196/jmir.6688] [Medline: 28526671]
21. Covvey H. Healthcare as a complex adaptive system. In: eHealth Research, Theory and Development. London: Routledge; 2018:89-110.
22. Noblit G, Hare R. Meta-Ethnography. Thousand Oaks: Sage Publications; 1988.
23. Booth A, Noyes J, Flemming K, Gerhardus A, Wahlster P, van der Wilt GJ. Guidance on choosing qualitative evidence synthesis methods for use in health technology assessments of complex interventions. 2016. URL: https://www.integrate-hta.eu/wp-content/uploads/2016/02/Guidance-on-choosing-qualitative-evidence-synthesis-methods-for-use-in-HTA-of-complex-interventions.pdf [accessed 2020-04-13]
24. Booth A, Sutton A, Papaioannou D. Systematic Approaches to a Successful Literature Review. 2nd Edition. Thousand Oaks: Sage Publications; 2016.
25. Cahill M, Robinson K, Pettigrew J, Galvin R, Stanley M. Qualitative synthesis: a guide to conducting a meta-ethnography. Brit J Occup Therapy 2018 Jan 09;81(3):129-137. [doi: 10.1177/0308022617745016]
26. Campbell R, Pound P, Morgan M, Daker-White G, Britten N, Pill R, et al. Evaluating meta-ethnography: systematic analysis and synthesis of qualitative research. Health Technol Assess 2011 Dec;15(43):1-164 [FREE Full text] [doi: 10.3310/hta15430] [Medline: 22176717]
27. France EF, Cunningham M, Ring N, Uny I, Duncan EAS, Jepson RG, et al. Improving reporting of meta-ethnography: the eMERGe reporting guidance. Psychooncology 2019 Mar;28(3):447-458. [doi: 10.1002/pon.4915] [Medline: 30644150]
28. Cruz-Martínez R, Noort PD, Asbjørnsen R, van Niekerk JM, Wentzel J, Sanderman R, et al. Frameworks, models, and theories used in electronic health research and development to support self-management of cardiovascular diseases through remote monitoring technologies: protocol for a metaethnography review. JMIR Res Protoc 2019 Jul 16;8(7):e13334 [FREE Full text] [doi: 10.2196/13334] [Medline: 31313659]
29. Stoll CRT, Izadi S, Fowler S, Green P, Suls J, Colditz GA. The value of a second reviewer for study selection in systematic reviews. Res Synth Methods 2019 Jul 04:1. [doi: 10.1002/jrsm.1369] [Medline: 31272125]
30. McDonagh M, Peterson K, Raina P, Chang S, Shekelle P. Agency for Healthcare Research and Quality (US). 2013. Methods guide for effectiveness and comparative effectiveness reviews: avoiding bias in selecting studies URL: https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/methods-guidance-bias_methods.pdf [accessed 2020-04-13]
31. Baker TB, Gustafson DH, Shaw B, Hawkins R, Pingree S, Roberts L, et al. Relevance of CONSORT reporting criteria for research on eHealth interventions. Patient Educ Couns 2010 Dec;81 Suppl:S77-S86 [FREE Full text] [doi: 10.1016/j.pec.2010.07.040] [Medline: 20843621]
32. Eysenbach G. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
33. Atkins S, Lewin S, Smith H, Engel M, Fretheim A, Volmink J. Conducting a meta-ethnography of qualitative literature: lessons learnt. BMC Med Res Methodol 2008 Apr 16;8:21 [FREE Full text] [doi: 10.1186/1471-2288-8-21] [Medline: 18416812]
34. Britten N, Pope C. Medicine taking for asthma: a worked example of meta-ethnography. In: Hannes K, Lockwood C, editors. Synthesizing Qualitative Research: Choosing the Right Approach. Hoboken: Wiley Blackwell; 2012.
35. Malpass A, Shaw A, Sharp D, Walter F, Feder G, Ridd M, et al. "Medication career" or "moral career"? The two sides of managing antidepressants: a meta-ethnography of patients' experience of antidepressants. Soc Sci Med 2009 Jan;68(1):154-168. [doi: 10.1016/j.socscimed.2008.09.068] [Medline: 19013702]
36. Tong A, Flemming K, McInnes E, Oliver S, Craig J. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. BMC Med Res Methodol 2012 Nov 27;12:181 [FREE Full text] [doi: 10.1186/1471-2288-12-181] [Medline: 23185978]
37. Toye F, Seers K, Allcock N. A meta-ethnography of patients' experience of chronic non-malignant musculoskeletal pain. Health Serv Deliv Res 2013 Nov 21;1(12):1. [doi: 10.3310/hsdr01120] [Medline: 25642538]
38. Toye F, Seers K, Allcock N, Briggs M, Carr E, Barker K. Meta-ethnography 25 years on: challenges and insights for synthesising a large number of qualitative studies. BMC Med Res Methodol 2014 Jun 21;14:80 [FREE Full text] [doi: 10.1186/1471-2288-14-80] [Medline: 24951054]
39. Gleich DF. PageRank beyond the web. SIAM Rev 2015 Jan;57(3):321-363. [doi: 10.1137/140976649]
40. Knutas A, Hajikhani A, Salminen J, Ikonen J, Porras J. Cloud-based bibliometric analysis service for systematic mapping studies. 2015 Presented at: Proceedings of the 16th International Conference on Computer Systems and Technologies; 2015; Dublin p. 184-191. [doi: 10.1145/2812428.2812442]
41. Blei D, Ng A, Jordan M. Latent Dirichlet allocation. J Mach Learn Res 2003;3:993-1022 [FREE Full text]
42. Badampudi D, Wohlin C, Petersen K. Experiences from using snowballing and database searches in systematic literature studies. 2015 Presented at: Proceedings of the 19th International Conference on Evaluation and Assessment in Software Engineering; 2015; Nanjing p. 1-10. [doi: 10.1145/2745802.2745818]

43. France EF, Uny I, Ring N, Turley RL, Maxwell M, Duncan EAS, et al. A methodological systematic review of meta-ethnography conduct to articulate the complex analytical phases. BMC Med Res Methodol 2019 Feb 18;19(1):35 [FREE Full text] [doi: 10.1186/s12874-019-0670-7] [Medline: 30777031]
44. 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]
45. Athilingam P, Labrador MA, Remo EFJ, Mack L, San JAB, Elliott AF. Features and usability assessment of a patient-centered mobile application (HeartMapp) for self-management of heart failure. Appl Nurs Res 2016 Nov;32:156-163. [doi: 10.1016/j.apnr.2016.07.001] [Medline: 27969021]
46. Athilingam P, Clochesy JM, Labrador MA. Intervention mapping approach in the design of an interactive mobile health application to improve self-care in heart failure. Comput Inform Nurs 2018 Feb;36(2):90-97. [doi: 10.1097/CIN.0000000000000383] [Medline: 28901967]
47. Athilingam P, Jenkins BA, Zumpano H, Labrador MA. Mobile technology to improve heart failure outcomes: a proof of concept paper. Appl Nurs Res 2018 Feb;39:26-33. [doi: 10.1016/j.apnr.2017.10.018] [Medline: 29422170]
48. Band R, Morton K, Stuart B, Raftery J, Bradbury K, Yao G, et al. Home and Online Management and Evaluation of Blood Pressure (HOME BP) digital intervention for self-management of uncontrolled, essential hypertension: a protocol for the randomised controlled HOME BP trial. BMJ Open 2016 Nov 07;6(11):e012684 [FREE Full text] [doi: 10.1136/bmjopen-2016-012684] [Medline: 27821598]
49. Band R, Bradbury K, Morton K, May C, Michie S, Mair FS, et al. Intervention planning for a digital intervention for self-management of hypertension: a theory-, evidence- and person-based approach. Implement Sci 2017 Feb 23;12(1):1-13 [FREE Full text] [doi: 10.1186/s13012-017-0553-4] [Medline: 28231840]
50. Bradbury K, Morton K, Band R, May C, McManus R, Little P, et al. Understanding how primary care practitioners perceive an online intervention for the management of hypertension. BMC Med Inform Decis Mak 2017 Jan 9;17(1):1-13. [doi: 10.1186/s12911-016-0397-x] [Medline: 120657722]
51. Rahimi K, Velardo C, Triantafyllidis A, Conrad N, Shah SA, Chantler T, SUPPORT-HF Investigators, et al. A user-centred home monitoring and self-management system for patients with heart failure: a multicentre cohort study. Eur Heart J Qual Care Clin Outcomes 2015 Nov 01;1(2):66-71 [FREE Full text] [doi: 10.1093/ehjqcco/qcv013] [Medline: 29474596]
52. Triantafyllidis A, Velardo C, Chantler T, Shah SA, Paton C, Khorshidi R, SUPPORT-HF Investigators. A personalised mobile-based home monitoring system for heart failure: the SUPPORT-HF Study. Int J Med Inform 2015 Oct;84(10):743-753. [doi: 10.1016/j.ijmedinf.2015.05.003] [Medline: 26037921]
53. Chantler T, Paton C, Velardo C, Triantafyllidis A, Shah SA, Stoppani E, et al. Creating connections—the development of a mobile-health monitoring system for heart failure: qualitative findings from a usability cohort study. Digit Health 2016;2:2055207616671461 [FREE Full text] [doi: 10.1177/2055207616671461] [Medline: 29942568]
54. Walsh D, Moran K, Cornelissen V, Buys R, Claes J, Zampognaro P, et al. The development and codesign of the PATHway intervention: a theory-driven eHealth platform for the self-management of cardiovascular disease. Transl Behav Med 2019 Jan 01;9(1):76-98. [doi: 10.1093/tbm/iby017] [Medline: 29554380]
55. Walsh DM, Moran K, Cornelissen V, Buys R, Cornelis N, Woods C. Electronic health physical activity behavior change intervention to self-manage cardiovascular disease: qualitative exploration of patient and health professional requirements. J Med Internet Res 2018 May 08;20(5):e163 [FREE Full text] [doi: 10.2196/jmir.9181] [Medline: 29739740]
56. Bartlett YK, Haywood A, Bentley CL, Parker J, Hawley MS, Mountain GA, et al. The SMART personalised self-management system for congestive heart failure: results of a realist evaluation. BMC Med Inform Decis Mak 2014 Nov 25;14(1):1. [doi: 10.1186/s12911-014-0109-3] [Medline: 25421307]
57. Duff O, Walsh D, Malone S, McDermott L, Furlong B, O'Connor N, et al. MedFit app, a behavior-changing, theoretically informed mobile app for patient self-management of cardiovascular disease: user-centered development. JMIR Form Res 2018 Apr 27;2(1):e8 [FREE Full text] [doi: 10.2196/formative.9550] [Medline: 30684426]
58. McGillicuddy J, Weiland A, Treiber E, Gregoski M, Patel S, Davidson L. Invited paper: Facilitating medication adherence and eliminating therapeutic inertia using wireless technology: proof of concept findings with uncontrolled hypertensives and kidney transplant recipients. WH12 Proc Conf Wireless Health 2012:1. [doi: 10.1145/2448096.2448108]
59. Srinivas P, Cornet V, Holden R. Human factors analysis, design, and evaluation of Engage, a consumer health IT application for geriatric heart failure self-care. Int J Hum Comput Interact 2017;33(4):298-312 [FREE Full text] [doi: 10.1080/10447318.2016.1265784] [Medline: 30429638]
60. Villalba E, Salvi D, Peinado I, Ottaviano M. Validation results of the user interaction in a heart failure management system. 2009 Presented at: 2009 International Conference on eHealth, Telemedicine, and Social Medicine; 2009; Cancun. [doi: 10.1109/etelemed.2009.41]
61. Baek H, Suh J, Kang S, Kang S, Lim TH, Hwang H, et al. Enhancing user experience through user study: design of an mhealth tool for self-management and care engagement of cardiovascular disease patients. JMIR Cardio 2018 Feb 09;2(1):e3 [FREE Full text] [doi: 10.2196/cardio.9000] [Medline: 31758783]
62. Centre for Evidence-Based Medicine. 2014. Study designs URL: https://www.cebm.net/2014/04/study-designs/ [accessed 2019-07-16]

63. Quesenbery W. Dimensions of usability: defining the conversation, driving the proces. 2003. URL: https://www.wqusability.com/articles/5es-upa2003.pdf [accessed 2020-04-13]
64. Hearn G, Foth M. Action research in the design of new media and ICT systems. In: Kwansah-Aidoo K, editor. Topical Issues in Communications and Media Research. New York: Nova Science; 2005:79-94.
65. Winter R, Munn-Giddings C. A Handbook for Action Research in Health and Social Care. London: Psychology Press; 2001.
66. Meso P, Jain R. Agile software development: adaptive systems principles and best practices. Inf Syst Manag 2006 Jun;23(3):19-30. [doi: 10.1201/1078.10580530/46108.23.3.20060601/93704.3]
67. Dumas J, Dumas J, Redish J. A Practical Guide to Usability Testing. Bristol: Intellect books; 1999.
68. Michie S, Atkins L, West R. The Behaviour Change Wheel: A Guide to Designing Interventions. London: Silverback Publishing; 2014.
69. Michie S, van Stralen M, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
70. Sreejesh S. Business Research Methods: An Applied Orientation. Cham: Springer Nature; 2014.
71. Whittaker R, Merry S, Dorey E, Maddison R. A development and evaluation process for mHealth interventions: examples from New Zealand. J Health Commun 2012;17 Suppl 1:11-21. [doi: 10.1080/10810730.2011.649103] [Medline: 22548594]
72. Cooper A. About Face 3: The Essentials of Interaction Design. Hoboken: John Wiley & Sons, Inc; 2007.
73. Villalba ME, Salvi D, Ottaviano M, Peinado I, Waldmeyer M. Iterative user interaction design for wearable and mobile solutions to assess cardiovascular chronic diseases. In: Human Computer Interaction. London: IntechOpen; 2008.
74. Bartholomew L. Intervention Mapping: Designing Theory and Evidence-Based Health Promotion Programs with PowerWeb. New York: McGraw-Hill Higher Education; 2001.
75. Sotirovski D. Heuristics for iterative software development. IEEE Softw 2001;18(3):66-73. [doi: 10.1109/52.922728]
76. Campbell NC, Murray E, Darbyshire J, Emery J, Farmer A, Griffiths F, et al. Designing and evaluating complex interventions to improve health care. BMJ 2007 Mar 03;334(7591):455-459 [FREE Full text] [doi: 10.1136/bmj.39108.379965.BE] [Medline: 17332585]
77. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
78. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. Int J Nurs Stud 2013 May;50(5):587-592. [doi: 10.1016/j.ijnurstu.2012.09.010] [Medline: 23159157]
79. Collins LM, Murphy SA, Nair VN, Strecher VJ. A strategy for optimizing and evaluating behavioral interventions. Ann Behav Med 2005 Aug;30(1):65-73. [doi: 10.1207/s15324796abm3001_8] [Medline: 16097907]
80. Valdez RS, Holden RJ, Novak LL, Veinot TC. Transforming consumer health informatics through a patient work framework: connecting patients to context. J Am Med Inform Assoc 2015 Jan;22(1):2-10. [doi: 10.1136/amiajnl-2014-002826] [Medline: 25125685]
81. Yardley L, Ainsworth B, Arden-Close E, Muller I. The person-based approach to enhancing the acceptability and feasibility of interventions. Pilot Feasibility Stud 2015;1:37 [FREE Full text] [doi: 10.1186/s40814-015-0033-z] [Medline: 27965815]
82. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
83. Marchal B, van Belle S, van Olmen J, Hoerée T, Kegels G. Is realist evaluation keeping its promise? A review of published empirical studies in the field of health systems research. Evaluation 2012 Apr 22;18(2):192-212. [doi: 10.1177/1356389012442444]
84. Pawson R, Tilley N, Tilley N. Realistic Evaluation. Thousand Oaks: Sage Publications; 1997.
85. Holden RJ, Carayon P, Gurses AP, Hoonakker P, Hundt AS, Ozok AA, et al. SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients. Ergonomics 2013;56(11):1669-1686 [FREE Full text] [doi: 10.1080/00140139.2013.838643] [Medline: 24088063]
86. Blank S, Dorf B. The Startup Owner's Manual: The Step-By-Step Guide for Building a Great Company. Uvalda: K&S Ranch; 2012.
87. Humayoun S, Dubinsky Y, Catarci T. A Three-Fold Integration Framework to Incorporate User-Centered Design into Agile Software Development. Berlin: Springer; 2011.
88. 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]
89. Monk A. User-centred design. In: International Conference on Home-Oriented Informatics and Telematics. Cham: Springer; 2000:181-190.

90. Holden R, Voida S, Savoy A, Jones J, Kulanthaivel A. Human factors engineering and human-computer interaction: supporting user performance and experience. In: Finnell JT, Dixon BE, editors. Clinical Informatics Study Guide: Text and Review. Cham: Springer International Publishing; 2016:287-307.
91. International Organization for Standardization. 1999. Human-centred design processes for interactive systems URL: https://www.iso.org/obp/ui/#iso:std:iso:13407:ed-1:v1:en [accessed 2020-04-13]
92. Daniels J, Fels S, Kushniruk A, Lim J, Ansermino JM. A framework for evaluating usability of clinical monitoring technology. J Clin Monit Comput 2007 Oct;21(5):323-330. [doi: 10.1007/s10877-007-9091-y] [Medline: 17701381]
93. Sweller J. Cognitive load during problem solvingffects on learning. Cognitive Sci 1988;12(2):257-285. [doi: 10.1207/s15516709cog1202_4]
94. Moreno R, Mayer RE. A cognitive theory of multimedia learning: implications for design principles. J Educ Psychol 1999;91(2):358-368.
95. Leventhal H, Phillips L, Burns E. The Common-Sense Model of Self-Regulation (CSM): a dynamic framework for understanding illness self-management. J Behav Med 2016 Dec;39(6):935-946. [doi: 10.1007/s10865-016-9782-2] [Medline: 27515801]
96. Carver CS, Scheier MF. Control theory: a useful conceptual framework for personality-social, clinical, and health psychology. Psychol Bull 1982 Jul;92(1):111-135. [Medline: 7134324]
97. Hirsch E, Silverstone R. Consuming Technologies: Media and Information in Domestic Spaces. London: Routledge; 2003.
98. Silverstone R, Haddon L. Design and the domestication of information and communication technologies: technical change and everyday life. In: Communication by Design: The Politics of Information and Communication Technologies. Oxford: Oxford University Press; 1996:44-74.
99. Fisher W, Fisher J, Harman J. The information-motivation-behavioraI skills model: A general social psychological approach to understanding and promoting health behavior. In: Social Psychological Foundations of Health and Illness. Malden: Blackwell Publishing; 2003:82-106.
100. Moreno R. Multimedia learning with animated pedagogical agents. In: The Cambridge Handbook of Multimedia Learning. Cambridge: Cambridge University Press; 2005:507-524.
101. May C. A rational model for assessing and evaluating complex interventions in health care. BMC Health Serv Res 2006;6:86 [FREE Full text] [doi: 10.1186/1472-6963-6-86] [Medline: 16827928]
102. May CR, Mair FS, Dowrick CF, Finch TL. Process evaluation for complex interventions in primary care: understanding trials using the normalization process model. BMC Fam Pract 2007 Jul 24;8(1):42. [doi: 10.1186/1471-2296-8-42]
103. Murray E, Treweek S, Pope C, MacFarlane A, Ballini L, Dowrick C, et al. Normalisation process theory: a framework for developing, evaluating and implementing complex interventions. BMC Med 2010;8:63 [FREE Full text] [doi: 10.1186/1741-7015-8-63] [Medline: 20961442]
104. Savery J, Duffy T. Problem based learning: An instructional model and its constructivist framework. Educ Technol 1995;35(5):31-38.
105. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
106. Ryan RM, Patrick H, Deci EL, Williams GC. Facilitating health behaviour change and its maintenance: interventions based on self-determination theory. Eur Health Psychol 2008;10(1):2-5.
107. Bandura A. Social Learning Theory: Group & Organization Studies. Englewood Cliffs: Prentice-Hall; 1977:384-385.
108. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs: Prentice-Hall, Inc; 1986.
109. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
110. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Manag Sci 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
111. Venkatesh V, Davis F, Morris M. Dead or alive? The development, trajectory and future of technology adoption research. J Assoc Inf Syst 2007 Apr;8(4):267-286. [doi: 10.17705/1jais.00120]
112. Venkatesh, Thong, Xu. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Quarterly 2012;36(1):157. [doi: 10.2307/41410412]
113. Carman KL, Dardess P, Maurer M, Sofaer S, Adams K, Bechtel C, et al. Patient and family engagement: a framework for understanding the elements and developing interventions and policies. Health Aff (Millwood) 2013 Feb;32(2):223-231 [FREE Full text] [doi: 10.1377/hlthaff.2012.1133] [Medline: 23381514]
114. Granja C, Janssen W, Johansen MA. Factors determining the success and failure of ehealth interventions: systematic review of the literature. J Med Internet Res 2018 May 01;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]
115. Greenhalgh T, A'Court C, Shaw S. Understanding heart failure; explaining telehealth—a hermeneutic systematic review. BMC Cardiovasc Disord 2017 Dec 14;17(1):156 [FREE Full text] [doi: 10.1186/s12872-017-0594-2] [Medline: 28615004]
116. Harvey J, Dopson S, McManus RJ, Powell J. Factors influencing the adoption of self-management solutions: an interpretive synthesis of the literature on stakeholder experiences. Implement Sci 2015 Nov 13;10:159 [FREE Full text] [doi: 10.1186/s13012-015-0350-x] [Medline: 26566623]

117. Borgnakke K. Meta-ethnography and systematic reviews—linked to the evidence movement and caught in a dilemma. Ethnography Educ 2016 Nov 09;12(2):194-210. [doi: 10.1080/17457823.2016.1253027]
118. Britten N, Campbell R, Pope C, Donovan J, Morgan M, Pill R. Using meta ethnography to synthesise qualitative research: a worked example. J Health Serv Res Policy 2002 Oct;7(4):209-215. [doi: 10.1258/135581902320432732] [Medline: 12425780]
119. Erasmus E. The use of street-level bureaucracy theory in health policy analysis in low- and middle-income countries: a meta-ethnographic synthesis. Health Policy Plan 2014 Dec;29 Suppl 3:70-78. [doi: 10.1093/heapol/czu112] [Medline: 25435538]
120. Morton K, Dennison L, May C, Murray E, Little P, McManus R, et al. Using digital interventions for self-management of chronic physical health conditions: a meta-ethnography review of published studies. Patient Educ Couns 2017 Apr;100(4):616-635. [doi: 10.1016/j.pec.2016.10.019] [Medline: 28029572]
121. Siau K, Long Y. Synthesizing e‐government stage models—a meta‐synthesis based on meta‐ethnography approach. Industr Manag Data Syst 2005 May;105(4):443-458. [doi: 10.1108/02635570510592352]
122. Noyes J, Booth A, Flemming K, Garside R, Harden A, Lewin S, et al. Cochrane Qualitative and Implementation Methods Group guidance series—paper 3: methods for assessing methodological limitations, data extraction and synthesis, and confidence in synthesized qualitative findings. J Clin Epidemiol 2018 May;97:49-58. [doi: 10.1016/j.jclinepi.2017.06.020] [Medline: 29247700]
123. Kok G, Peters LWH, Ruiter RAC. Planning theory- and evidence-based behavior change interventions: a conceptual review of the intervention mapping protocol. Psicol Reflex Crit 2017 Oct 18;30(1):19 [FREE Full text] [doi: 10.1186/s41155-017-0072-x] [Medline: 32026109]
124. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
125. Hsu C, Sandford B. The Delphi technique: making sense of consensus. Pract Assess Res Eval 2007;12(10):1-8. [doi: 10.7275/pdz9-th90]

Abbreviations

CeHRes: Center for eHealth Research
CHF PSMS: Congestive Heart Failure—Personalized Self-Management System
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
CVD: cardiovascular diseases
eHealth: electronic health
eMERGe: Meta-ethnography Reporting Guidance
HOME BP: Home and Online Management and Evaluation of Blood Pressure
MRC: Medical Research Council
PRISMA: Preferred Reporting Items for Systematic Review and Meta-Analyses
SMASH: Smartphone Medication Adherence Stops Hypertension
SUPPORT HF: Seamless User-Centered Proactive Provision of Risk-Stratified Treatment for Heart Failure

Edited by G Eysenbach; submitted 06.09.19; peer-reviewed by K Deldar, M Johansson, D Amiri, J Huckins, M Kolotylo-Kulkarni; comments to author 02.12.19; revised version received 19.12.19; accepted 07.02.20; published 21.05.20.

Please cite as:
Cruz-Martínez RR, Wentzel J, Asbjørnsen RA, Noort PD, van Niekerk JM, Sanderman R, van Gemert-Pijnen JEWC
Supporting Self-Management of Cardiovascular Diseases Through Remote Monitoring Technologies: Metaethnography Review of Frameworks, Models, and Theories Used in Research and Development
J Med Internet Res 2020;22(5):e16157
URL: http://www.jmir.org/2020/5/e16157/
doi:10.2196/16157
PMID:32436852

Review

Ethical, Legal, and Social Issues Related to the Inclusion of Individuals With Intellectual Disabilities in Electronic Health Record Research: Scoping Review

Melissa Raspa[1], BA, MA, PhD; Rebecca Moultrie[1], AA; Laura Wagner[1], MPH; Anne Edwards[1], BA; Sara Andrews[1], MA, MPH; Mary Katherine Frisch[2], BA; Lauren Turner-Brown[2], PhD; Anne Wheeler[1], PhD

[1]RTI International, Research Triangle Park, NC, United States

[2]The University of North Carolina at Chapel Hill, TEACCH Autism Program, Chapel Hill, NC, United States

Corresponding Author:
Melissa Raspa, BA, MA, PhD
RTI International
3040 East Cornwallis Road
PO Box 12194
Research Triangle Park, NC, 27709
United States
Phone: 1 19195418736
Email: mraspa@rti.org

Abstract

Background: Data from electronic health records (EHRs) are increasingly used in the field of genetic research to further precision medicine initiatives. However, many of these efforts exclude individuals with intellectual disabilities, which often stem from genetic conditions. To include this important subpopulation in EHR research, important ethical, legal, and social issues should be considered.

Objective: The goal of this study was to review prior research to better understand what ethical, legal, and social issues may need further investigation when considering the research use of EHRs for individuals with genetic conditions that may result in intellectual disability. This information will be valuable in developing methods and best practices for involving this group in research given they are considered a vulnerable population that may need special research protections.

Methods: We conducted a scoping review to examine issues related to the use of EHRs for research purposes and those more broadly associated with genetic research. The initial search yielded a total of 460 unique citations. We used an evaluative coding process to determine relevancy for inclusion.

Results: This approach resulted in 59 articles in the following areas: informed consent, privacy and security, return of results, and vulnerable populations. The review included several models of garnering informed consent in EHR or genetic research, including tiered or categorical, blanket or general, open, and opt-out models. Second, studies reported on patients' concerns regarding the privacy and security of EHR or genetic data, such as who has access, type of data use in research, identifiability, and risks associated with privacy breach. The literature on return of research results using biospecimens examined the dissension in the field, particularly when sharing individualized genetic results. Finally, work involving vulnerable populations highlighted special considerations when conducting EHR or genetic research.

Conclusions: The results frame important questions for researchers to consider when designing EHR studies, which include individuals with intellectual disabilities, including appropriate safeguards and protections.

(J Med Internet Res 2020;22(5):e16734) doi:10.2196/16734

KEYWORDS

electronic health records; privacy; informed consent; intellectual disability; genetics

Introduction

Background

The field of genetics has grown exponentially over the last decade. Advances in whole genome and exome sequencing have made the diagnosis of genetic conditions easier than ever before. Conditions diagnosed through these approaches run the gamut from cancer to rare diseases [1-4]. Genetic testing also is increasingly being integrated into clinical care and viewed as a common practice to detect genetic causes of conditions [5]. Another mechanism for genetic discovery lies with the increasing use of electronic health records (EHRs). EHRs have become a more commonplace tool in genetics, allowing researchers to conduct studies on specific genetic conditions, including observational, epidemiological, descriptive, and comparative effectiveness studies, among others [6,7]. EHRs are also being used to recruit participants for research studies and clinical trials [8].

Adding to their utility, EHRs can be linked with biobank data to answer genotype and phenotype research questions [9]. For example, the Electronic Medical Records and Genomics (eMERGE) network, which began in 2007, uses biorepositories linked to EHRs to conduct electronic phenotyping to identify patterns and diagnose disease [10-12]. More recently, the movement toward precision medicine has inspired other similar initiatives, such as Geisinger's MyCode Community Health Initiative [13] and *All of Us*, sponsored by the National Institutes of Health [14]. Other work has focused on phenome-wide association studies that analyze different phenotypes to determine a genetic variant [15].

All of these programs are designed with the promise of providing tailored health interventions based on an individual's genetic makeup. Ultimately, this may lead to a reduction in health disparities but only if a diverse array of individuals is included [16]. With this in mind, there have been calls to expand enrollment to underrepresented individuals, including those with intellectual disabilities [17,18]. Given that many conditions that results in intellectual disability can have genetic underpinnings, the inclusion of this subgroup in EHR research can provide insights into biological causes to various diseases and comorbidities [17].

Objectives

Previous work has examined ethical, legal, and social issues related to EHR use as well as combining EHRs with genetic data. However, research conducted to date has focused on typically developing individuals (ie, individuals with no known or suspected genetic conditions) who wanted to contribute their data to advance science and clinical practice [19], and thus, has predominantly excluded individuals with specific genetic conditions or those with intellectual disabilities. The inclusion of this group of individuals achieves the goal of making research, in particular the precision medicine initiative, more representative and patient-centered, and also aligns with the desires of those with disabilities [20]. Thus, there is a need to further explore the barriers or challenges of including this group of individuals in EHR research.

The goal of this study was to review existing studies to better understand what ethical, legal, and social issues may need further investigation when considering the research use of EHRs for individuals with genetic conditions that may result in intellectual disability. This information will be valuable in developing methods and best practices for involving this group in future research.

Methods

Design

We conducted a scoping review of the literature to examine ethical, legal, and social issues related to the use of EHRs for research purposes or issues more broadly associated with genetic research. We chose a scoping review approach given the breadth of literature in this area. This also enabled us to synthesize information from diverse sources, including theoretical and narrative reviews, qualitative studies, and quantitative research. We first identified our search terms; conducted a search of relevant literature; reviewed, charted, and collated the information; and summarized key findings [21,22].

Search Terms

Given that we were interested in ethical, legal, and social issues related to the inclusion of individuals with genetic conditions that resulted in intellectual disability in EHR research, we conducted searches for the following two topic areas: (1) research use of EHRs and (2) ethical, legal, and social issues of genetics research. The research team developed search terms for each topic area **(**Textbox 1**)**. Inclusion criteria included peer review articles published in English between 2000 and 2018. We conducted our search using the following databases: PubMed (Medical Literature Analysis and Retrieval System Online, MEDLINE), Cumulative Index to Nursing and Allied Health Literature (CINAHL), Cochrane Review, PsycINFO, and Web of Science.

Textbox 1. Search terms by topic area.

Research use of electronic health records

- medical record, health record, electronic record, electronic health information, health data, clinical record, clinical data, Health Insurance Portability and Accountability Act, HIPAA, PHR, EMR, EHR, eHealth, or e-Health, and
- outcome, research, measur*, assess*, evaluat*, analy*, study, or studies, and
- cognitive decline, cognitive impairment*, intellectual disabilit*, developmental disabilit*, autism, fragile X, Alzheimer*, genetic*, genomic*, vulnerable population, cognitively impaired, intellectually disabled, developmentally disabled, autistic,

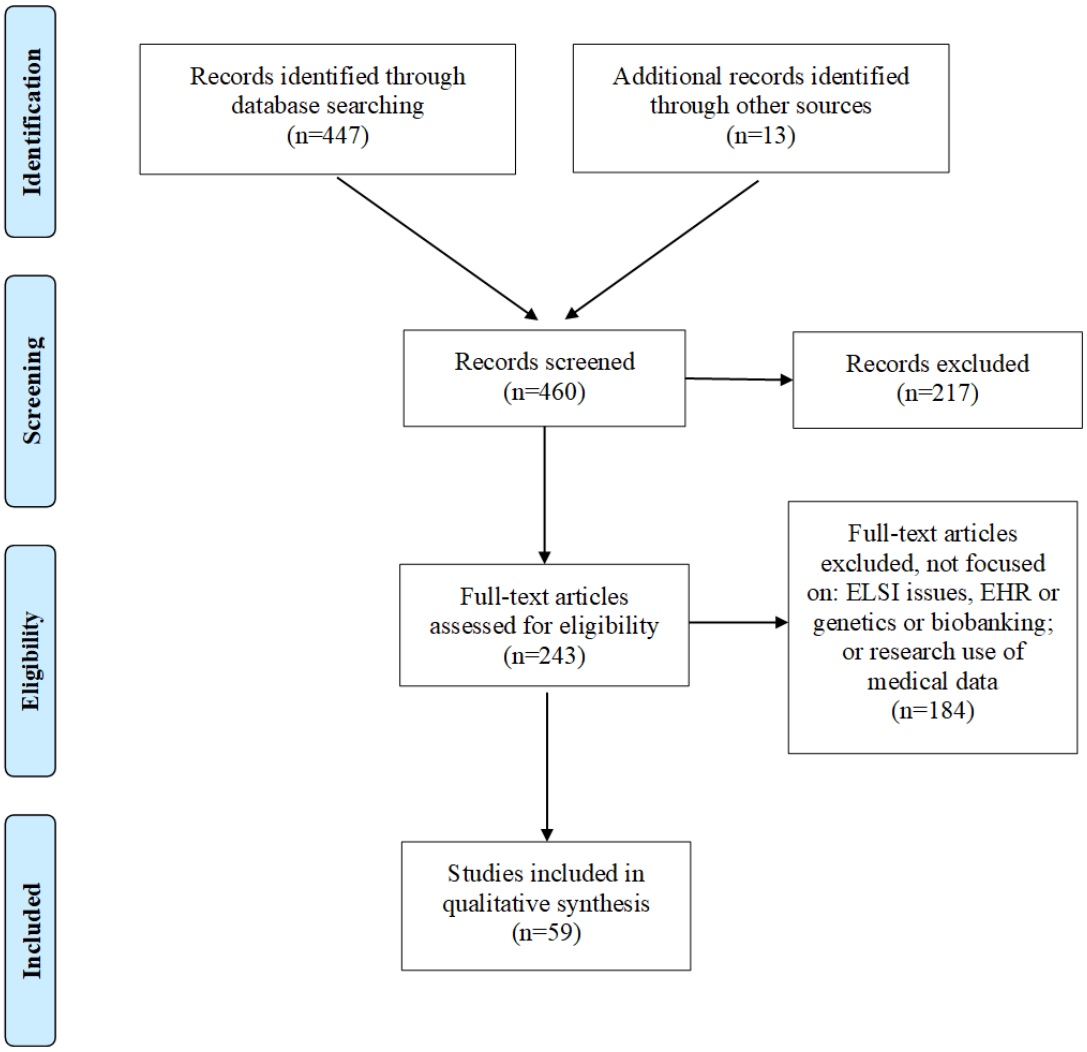
Ethical, legal, and social implications of genetics research

- ethical issue, legal issue, social issue, ethical implication, ethical requirement, guardian, ethic, or social, and
- consent issue, informed consent, consent, medical record, health record, electronic record, electronic medical record, electronic health record, electronic health information, health data, clinical record, clinical data, Health Insurance Portability and Accountability Act, HIPAA, PHR, EMR, EHR, eHealth, or e-Health, and
- cognitive decline, cognitive impairment, intellectual disabilit*, developmental disabilit*, autism, fragile X, Alzheimer*, genetic*, genomic*, cognitively impaired, intellectually disabled, developmentally disabled, autistic

Abstract Review

The initial search yielded a total of 447 unique citations (duplicates removed): 206 in the research use of EHRs topic area and 241 in the ethical, legal, and social issues of genetics research topic area. We identified an additional 13 articles outside our initial search parameters, mainly from ethics journals that were not indexed in MEDLINE. The study team developed a coding structure in which each article was scored on a 4-point scale, with 0 being not at all relevant and 3 being extremely relevant. A pair of researchers were assigned the same abstracts and conducted blind coding of each to determine relevance for inclusion and whether a full-text review of the article was appropriate. The pair met to review and compare scores. In cases where scores did not match, researchers discussed and arrived at a consensus score. In total, 243 abstracts were deemed relevant (scored a 2 or 3) to obtain the full text. This included 120 in the research use of EHRs category and 123 in the ethical, legal, and social implications of genetics research category. The full-text articles were then reviewed, key findings were extracted and charted, and themes were categorized. The scoping review culminated in a total of 62 articles: 26 on the research use of EHRs and 36 on the ethical, legal, and social issues of genetics research. These full-text articles were reviewed, and themes were extracted (see Figure 1).

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart. EHR: electronic health record; ELSI: ethical, legal, and social issues.

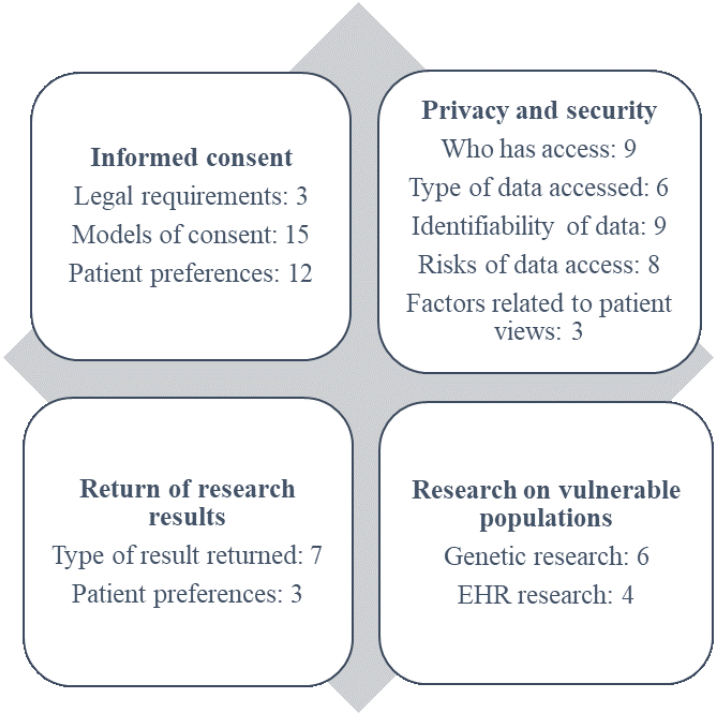


Results

Overview

We grouped the results into four broad ethical, legal, and social issues related to EHRs and genetic research that would applicable to those with intellectual disabilities (see Figure 2). The first section covers issues related to informed consent for the use of EHRs in research. This section includes a review of the legal requirements in the United States and also describes possible models of obtaining informed consent for EHRs and genetics research. The next section provides an overview of issues related to the privacy and security of EHRs and genetics research, including studies that examined preferences for who has access to information in the health record and what information is accessed. The third section examines the return of research results, which is most often considered within the context of genetics and biobank research. The final section discusses unique considerations for conducting research with vulnerable populations.

Figure 2. Number of articles included by ethical, legal, and social issue. EHR: electronic health record.



Informed Consent

Legal Requirements of Informed Consent

The first issue faced when conducting research with EHRs or on genetic conditions is whether informed consent is needed [23,24]. As set forth in the Common Rule and specified in US regulatory requirements, anyone conducting research with human subjects must have an institutional review board (IRB) review their study before initiation. Under the new federal regulations, referred to as the Final Rule [25], the definition of research remains unchanged as "a systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge" (see 45CFR46.102(l)). However, the classification of "human subjects" now includes detailed definitions about identifiable private information (ie, "private information for which the identity of the subject is or may readily be ascertained by the investigator or associated with the information") and identifiable biospecimens (ie, "a biospecimen for which the identity of the subject is or may readily be ascertained by the investigator or associated with the biospecimen;" see 45CFR46.102(e)).

If informed consent is required, it can be obtained directly from the research participant or the participant's legally authorized representative. A legally authorized representative is determined by law or court authority. Under the Final Rule, individuals with intellectual disabilities or those with genetic conditions are not specifically referenced. However, research participants who may be "vulnerable to coercion or undue influence" or those with "impaired decision-making ability" are mentioned. Examples of these potential participants may include children, prisoners, or those with intellectual disabilities.

In the informed consent form, researchers need to include whether any identifiable private information will be used as part of the study. It is unclear if a specific genetic condition, intellectual disability would fall under the definition of private identifiable information within the Final Rule. The code simply mentions that definition of will be reviewed regularly and updated as needed. Until then it appears that agencies who implement this policy will need to define what falls under the umbrella of identifiable private information. The Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule (see 45CFR160-164), which covers EHRs, does define genetic information and considers it protected health information. Like the Final Rule, HIPAA has specific requirements about sharing information, including data from an individual's EHR. Thus, EHR research on specific genetic conditions would need to be reviewed by an IRB to determine whether informed consent should be obtained from participants.

There are cases, however, when a waiver of informed consent can be granted. The following criteria must be met to request a waiver: (1) the research involves no more than minimal risk to the subjects; (2) the research could not practicably be carried out without the requested waiver or alteration; (3) if the research involves using identifiable private information or identifiable biospecimens, the research could not practicably be carried out without using such information or biospecimens in an identifiable format; (4) the waiver or alteration will not adversely affect the rights and welfare of the subjects; and (5) whenever appropriate, the subjects or legally authorized representatives will be provided with additional pertinent information after participation (see 45CFR46.116(f)). In general, local IRBs need to interpret the regulations and determine if a study qualifies as research on human subjects, and if so, whether a waiver of consent can be granted when conducting EHR research on genetic conditions.

When researchers meet all the requirements of a waiver of informed consent, it is often under the auspices of conducting research for the common good [26]. Examples include research

aimed at improving public health or the quality of medical care. However, EHR research without consent may violate the ethical principles of autonomy in the hopes of achieving beneficence and justice [27]. However, many bioethicists have argued in favor of EHR research without consent as long as the aforementioned requirements of a waiver are met, and proper privacy and security restrictions are in place [26,28,29].

Models of Informed Consent

If informed consent is not waived, several different models have been proposed for meeting human subject protection requirements and minimizing researcher burden. A tiered, or categorical, consent model that offers research participants options for how and when their data can be used has been espoused as best practice in the field on genetics research [30,31]. A tiered consent model allows participants to choose from many options, such as allowing their data to be used for the current study only, the current study as well as future studies about a specific condition or type of research, or unlimited use of their data [23,32,33]. This type of consent model could also include an option for reconsent, such as when a new subcategory of a genotype or phenotype for a particular condition is discovered that may need an additional level of consent before further research is conducted [34]. Although this approach is in line with the Belmont principle of *respect for persons* and encourages autonomy among research participants, some have criticized it given that it may be logistically challenging to implement and could potentially diminish the utility and applicability of clinical data, whether information in EHRs or biospecimens, for future research purposes [35,36]. A recent evolution of the tiered consent model is dynamic consent in which researchers and participants have ongoing communication that allows participants to update their consent preferences over time and enables researchers to return results to those interested [37].

A second approach to informed consent is the use of a blanket, or general, consent model [12,23,34]. Blanket consent can be described as an all-or-nothing option for using EHR data or biospecimens in future research. In other words, participants are given the choice to consent to all future research use of their data without being given the details of what that research may entail. The National Institutes of Health now asks funded investigators to include broad data sharing as part of their informed consent forms for genetic research studies [38]. Similar to the tiered consent model, there are advocates and critics of this approach. Proponents cite that it covers all the necessary ethical and legal requirements of informed consent, whereas detractors argue that the lack of specifics on future use means that it is more akin to providing permission than consent that is truly informed [39,40].

The open consent model, used by the Personal Genome Project, takes the blanket consent model a step further by stating up front that participants' data may be accessed broadly and there are no guarantees of anonymity, privacy, or confidentiality [34]. The goal of this approach is to provide more specifics about the possible ways data may be used and the risks of participating. Advocates of this approach argue that being open and honest about the study design and potential uses of data will promote trust and willingness to participate.

A final model of informed consent that has been proposed involves opting out of research participation. The opt-out model requires a participant to actively withdraw from the research. If they do not withdraw, it is assumed they passively consent to participate. This model was adopted by the Iceland biobank project, a national effort to merge three sources of data—EHR, genealogical, and genetic—into one database for research purposes [23]. The opt-out model, however, was not consistently applied across the data sources and was met with criticism by those who felt it did not offer adequate control for research participants or meet ethical and legal requirements [32]. It seems there is no uniform approach to seeking consent for biobank research [41].

Few studies have focused on informed consent models for accessing only EHR data, without a link to biospecimens. In a commentary on whether it is ethical to access EHR data, even identifiable information, without consent, Miller asserts the answer lies in the balance between risks and benefits [26]. In cases where the public good outweighs the personal risks to an individual, then informed consent can be waived under the auspices of a utilitarian philosophy. However, the objection to this approach is that it violates an individual's right to privacy.

Ultimately, under the Final Rule and HIPAA Privacy Rule, this decision rests in the hands of IRBs. However, in a study of IRB chairs and administrators that were presented with a hypothetical EHR research protocol, opinions varied widely as to whether informed consent was needed, even when identifying information was not included [42]. IRB chairs and administrators who stated that consent was needed reasoned that there was a risk for reidentification. Even though identifiable information (eg, last name) was not being extracted, other information such as birth date, postal code, and ethnic origin were included. IRB chairs and administrators who said consent was not needed stated that the data were anonymous or that the identifiability of participants was not relevant to determining if consent was required.

Patient Preferences Related to Informed Consent

Several US-based studies have examined the preferences of patients related to providing informed consent for use of clinical data, including both biospecimens and EHRs. In one large survey study, most patients were willing to participate in a biobank [43]. However, some studies suggest that participants tend to prefer a broad consent model, as long as no personally identifiable information is used [44,45]. In another study, public sharing of deidentified data was favored by the majority of participants when either a tiered or blanket option was given [46]. Sharing of data has not been shown to be related to level of understanding or recall of the informed consent process [47]. However, preferences for models of consent have been shown to vary by some sociodemographic factors, such as race, ethnicity, income, and education [43,45,46].

Consent preferences are also driven by what information is being accessed and how it is being used. Patients favored seeking consent for the use of medical history data and

treatment-related information [48]. Reconsent was preferred when researchers were investigating a health condition that was unrelated to the one for which they originally gave consent, if deidentified data were shared with an investigator at a different institution, and when a child for whom a parent had previously given consent had reached adulthood [43,49]. Patients' decision to consent to research use of their biospecimen is also related to personal characteristics, including whether a participant has an existing genetic condition [49].

Studies in Europe and Canada have found similar preferences. Surveys and focus groups with patients in Ireland and the United Kingdom indicated a preference for some level of control over use of EHR data for research, with many favoring a study-by-study or tiered consent model [50-52]. In an Italian study, over half of those enrolled in a national twin registry were against the use of their EHR data without informed consent [53]. Similarly, in Canada, participants preferred to be asked for consent either verbally or in writing before releasing their EHR data to researchers [54].

Privacy and Security

Who Has Access to Data for Research Use

One of the most common ethical, legal, and social issues related to using EHRs for research is who has access [27,33,55]. Patients often had the most concern about health insurance and pharmaceutical companies as well as government agencies using their EHR data for research purposes [44,54,56-58]. Pereira et al [59] found that 66.9% (328/490), 44.9% (220/490), and 40.0% (196/490), respectively, did not want insurance companies, pharmaceutical companies, and the government to have access to their health information. Patients worried that research conducted by pharmaceutical companies would be used to promote products, and there was potential for insurance companies to deny coverage for patients [54]. These companies were seen as less "legitimate" to patients because the research they were likely to conduct was not "pure science," which aimed to generate knowledge for the common good [58]. Patients were also concerned with unauthorized access to their medical information or sharing of EHR data with noncredible researchers [57]. One example cited was government researchers accessing EHR data but then selling it to a pharmaceutical company [56]. Therefore, patients often desired greater oversight of individuals who conducted EHR research within these organizations [44]. Typically, the amount of trust a patient had in the researcher or organization conducting the research was related to the patient's views on privacy [44,51,54,56].

What Type of Data Are Accessed for Research

Some patients reported that they would prefer to limit access to certain information in their medical records to maintain privacy [58,60]. Patients were less likely to share sensitive information such as substance abuse history, mental health information, sexual health information, domestic violence records, reproductive health records, and genetic information and would want to be asked permission before sharing this information with researchers [60]. In particular, patients were concerned with the disclosure of conditions viewed as stigmatizing, such as HIV/AIDS and mental illness, to researchers [56-58]. One study reported that patients viewed reproductive and mental health information as more sensitive than genetic data [58]. To ensure privacy, some patients wanted to limit access to portions of their EHR to their health care provider [61] or not include any sensitive information in their EHR [16,33].

Identifiability of Data

Despite widespread support for research use of EHR data and biospecimens, patients often preferred that their information remain anonymous [49,51]. Hull [62] conducted interviews with patients of academic medical centers to examine attitudes and preferences on the use of anonymous versus identifiable samples for genetic research. On the basis of responses to two hypothetical scenarios, 72.96% (850/1165) preferred to be informed about the use of their anonymous blood samples, whereas 83.35% (971/1165) of patients wanted to be told if it was an identifiable sample. However, only 23.38% (271/1159) of patients responded differently to the two scenarios, with most (196/1159, 16.91%) indicating they wanted to be notified about the use of an identifiable sample and did not need to know if it was done with an anonymous sample.

In biobanking research, similar privacy concerns have been raised about lack of oversight and the potential for biospecimens to be identified [63,64]. Although technology has enabled large-scale genomic studies to be conducted, the sharing of data with outside researchers and linking of biospecimens with EHRs has increased the potential for reidentification of participants [23,63]. Improved data mining technology has also contributed to privacy and identifiability concerns [65]. Despite the fact that some research has shown that reidentification of biospecimens and EHR data is difficult when proper security controls are in place [66,67], concerns about privacy and security are likely to be an ongoing issue.

Risks Associated With Breach of Privacy

Patients voiced concerns about the implications of breaches of privacy. Common fears were discrimination, stigmatization, or psychosocial discomfort if medical information was shared with insurance companies or employers [22,23,51,63]. Similarly, employment and life insurance discrimination concerns were raised by community advisory group members regarding the presence of genomic information in the medical record [68]. In addition, concerns were expressed by some that genetic health information may affect their medical care—specifically the discontinuation or withholding of treatment [68]. In genetics research, patients also cited worries about release of information to family members without their consent [22,63]. Moreover, genetic information about risk for having a late-onset disease, such as Huntington's, could lead to social stigmatization as well as insurance and employment discrimination [22,65,69].

Factors Related to Patients' Views on Privacy

Three studies identified some individual factors related to views on privacy. Older people, described as individuals aged 50 years and older, reported less concern with privacy of genetic information than did younger people in a focus group study [58]. Gender differences were found related to the potential consequences of a security breach. Male patients reported

concern about impacts on employment, finances, and insurance, whereas female patients were concerned about risks related to social discomfort and embarrassment [51]. Finally, those who had more experience with using computers were less concerned about security issues [59].

Return of Research Results

Return of Individualized or Aggregate Results

Bioethicists have debated whether to return the results of genetic research to biobank participants. Much of the discussion has focused on the return of individualized versus aggregate results. Most agree that returning aggregate or summary results, through a newsletter or website, is appropriate. This is seen as both a measure of accountability for the researchers conducting the study and a sign of respect for the participants [23]. The provision of individual genetic results from a research study, though, has been a more contentious subject. Ethical arguments in favor of returning results are grounded in the principle of *respect for persons* and *beneficence* [23]. Having access to genetic information about oneself may have an impact on one's health, quality of life, and future decision making. Returning individualized results to participants also prevents researchers from being gatekeepers of important information. However, those against sharing individual results cite that research and clinical care should not be blurred, and an investigator providing genetic results is acting more like a clinician than a researcher. In addition, those opposed state that research and clinical testing are quite different, with the goal of the latter being to provide better care to a patient, whereas the former is concerned with creating generalizable knowledge [23].

Given this debate, several ethics boards and research committees have created recommendations or guidelines for when to share individualized results, including for which genetic conditions [31,70,71]. The eMERGE network has also addressed this issue [72]. Among these guidelines, there is consensus that individual results should only be returned if they are medically actionable, even if they were incidental findings. Researchers are also encouraged to thoughtfully consider the language in the informed consent form that details what types of results will be shared with participants [22]. However, in a recent study that examined informed consent forms, the majority of consent forms did not discuss return of results [73]. The authors acknowledge, though, that many of these forms may have been created years before any guidance was offered on this issue.

Patient Preference for Return of Research Results

When patients have been asked about their preferences for the return of research results, most indicate they want to know. In a survey of patients enrolled in a Health Maintenance Organization (HMO) plan, about two-thirds said they would be willing to participate in research within the HMO if written information about the study was provided to them and almost three-fourths if results were provided [74]. Another study asked participants about their views regarding collection of tissue samples and genetic testing [49]. Return of test results was strongly supported by all participants, with the highest level of support among those who were concerned about having a health condition that may be genetic. In a survey of preferences for return of results using donated tissue samples, just under half (129/271, 47.6%) reported they always wanted to be given individual results, 25.1% (68/271) wanted results only after the researcher had assessed the risks and benefits of sharing the information, and 27.3% (74/271) wanted results if their doctor thought the information would inform decision making related to their care [75]. Respondents who thought genetic information was more sensitive than other types of health information were more likely to want to be told all individual results.

Research on Vulnerable Populations

Genetic Research on Vulnerable Populations

Important ethical, legal, and social issues are raised when vulnerable populations are involved in genetic research [76]. Although similar to the issues already discussed, concerns related to recruitment, informed consent, privacy and security, and disclosure of results are often amplified when conducting research with vulnerable populations [23,77].

A clear example is when children are recruited to be part of genetic studies. One study asked adults whether they would want to be reconsented if their parents enrolled them in a biobank study when they were children. Overall, most respondents (799/1186, 67.37%) said they would be willing to continue participation, but almost half (543/1186, 45.78%) wanted to be reconsented [78]. In a review of procedures used in 6 different birth cohort studies, investigators found that blanket consent was never used. However, the studies varied in their approach to returning research results; although some provided participants and their families with routine clinical information and test results, other studies only informed participants if test results were abnormal for a treatable condition. Finally, all studies recontacted participants when they reached adulthood and allowed them the opportunity to withdraw or reconsent [79]. A recent study, though, examined an alternative consent model. When presented with an opt-out model of enrollment into a study that linked the child's EHR with biospecimens, most parents were supportive [80]. Parents stated there was little risk given that the data would be deidentified. Although some were concerned about the security of EHR data, none raised reservations about access to genetic information.

Electronic Health Record Research on Vulnerable Populations

Although studies have been conducted using EHRs of vulnerable populations, such as epidemiological research [81-83], very few have examined the ethical, legal, and social issues related to this type of research. Simon and colleagues [84] discussed the risks and benefits of conducting research on individuals with psychiatric conditions using large medical databases. Although privacy risks are of critical concern, the authors argued that these population-based research methods have provided many benefits, including information on prevalence estimates, treatment effectiveness, and impact of health policy. With appropriate safeguards, this research can also lead to reducing stigmatization and discrimination through education and outreach efforts based on research findings.

Discussion

Summary of Findings

In this study, we sought to better understand the ethical, legal, and social issues related to the use of EHRs of individuals with intellectual disabilities by examining similar issues, as well as patients' preferences, that have been raised in EHR research among the general public and in genetics or biospecimen research. The main themes we highlighted were related to obtaining informed consent, privacy and security of data, and the return of research results. With the push to expand precision medicine initiatives and EHR research to a diverse group of participants, it is important for researchers to consider the implications of these issues for individuals with genetic conditions, particularly those that result in intellectual disabilities. Below, we pose questions about each of these issues that researchers should consider when designing methods to involve those with intellectual disabilities in EHR research.

Given the issues related to informed consent that have been raised in prior studies, what are the best models of informed consent for use with individuals with genetic conditions who have intellectual disabilities? Is a broad or blanket consent model appropriate for this population? Would an opt-out model be ethically acceptable? or is the tiered consent model the only viable option? and would the type of consent model applicable for adults also be applicable for children with genetic conditions that result in intellectual disability? To answer these important questions, researchers should consider the decisional capacity of the individuals being asked to participate in the study. For example, individuals with fragile X syndrome, the most common inherited form of intellectual disability, may present with a range of impairment, from normal or mild delays to severe intellectual disability. Young children with a genetic condition, no matter their level of functioning, will require permission from a parent or guardian to participate in EHR research. However, when children with fragile X syndrome or other genetic conditions that result in intellectual disability reach adulthood, their capacity to consent should be assessed to determine if they adequately understand the study, including the risks and benefits of participating.

A large body of research has examined the capacity to consent in individuals with Alzheimer or other forms of cognitive impairment [85-88], and a handful have examined decisional capacity in individuals with intellectual disability [89-91]. This literature, combined with the ethical, legal, and social issues raised in this study, will help to inform the proper consent models for EHR researchers interested in recruiting participants who have a genetic condition that may result in intellectual disability. Experts agree, though, that a shared decision-making process should be employed to support these individuals to make informed consent choices [92-94].

Another set of questions arise regarding the privacy and security of EHR research conducted with individuals with intellectual disability that stems from a known or suspected genetic condition. Will these individuals have the same concerns as the general public regarding privacy and security of their EHR data, including who has access and what information researchers can see? Will these individuals be more reluctant to participate in EHR research due to the risk of identifiability, even if data have been deidentified? Or will they be less worried given that they may have more to benefit from EHR research? Much of the potential vulnerabilities of individuals with intellectual disabilities is related to the fear of coercion or possible stigmatization or discrimination that could result from misuse of research data [95].

EHR research conducted by pharmaceutical companies may trigger either support or apprehension from individuals with intellectual disability. Pharmaceutical companies play important roles in partnering with researchers to develop and test treatments; however, pharmaceutical companies may also be viewed unfavorably given the high cost of orphan drugs for some rare genetic conditions [95-98]. Similarly, those with an intellectual disability may be more or less willing to share sensitive or identifiable data with EHR researchers. If an individual has already been diagnosed with a genetic condition that results in intellectual disability, the risk of identification may be less significant if the individual's diagnosis is commonly known. But, for those with very rare genetic diseases, there may be a heightened risk of identifiability, mainly because there are so few patients with a given condition [99]. It is difficult to answer these questions because few studies have examined the preferences of individuals with intellectual disabilities about participating in research more broadly [100], and none have asked about EHR research in particular. More work is needed to understand the implications of privacy and security issues for individuals with intellectual disabilities to help researchers determine how best to address them. At a minimum, the risks and benefits related to privacy and security should be carefully outlined in the informed consent form so individuals with intellectual disabilities and their parents or legally authorized representative can weigh the risks and benefits of participation and make informed decisions. Ideally, EHR research and informed consent forms will continue to evolve as laws and other privacy safeguards for the use of clinical data are updated [101].

The central question about the return of EHR research results for individuals with intellectual disabilities is whether individual or aggregate results should be provided to participants. The same potential for the blurring of the lines between clinical care and research applies here as it does for those in the general public who join biobanks as "healthy" individuals. If a secondary condition related to a genetic condition is discovered through EHR research, is it necessary to notify the participant? The answer may lie in what is contained in the informed consent form. If the secondary finding is medically actionable and participants agree to be informed, then results should be returned. But, if the result does not require immediate medical attention, then researchers may not necessarily need to disclose it. Researchers should carefully consider the types of information they may learn from EHR research involving individuals with intellectual disabilities and determine *a priori* what types of results would qualify as needing to be returned. Then, participants should be provided with a choice as to whether they wish to be notified. If individual results are not returned, then

aggregate results in the form of patient-friendly summary reports or newsletters should be considered [102].

Limitations and Future Research

It is important to note there were limitations in our approach. First, although we were broad in our search terms and criteria, we may have missed pertinent articles that should have been included. Similarly, because this was a scoping review and not a systematic review, there may be other ethical, legal, or social issues related to EHR research on individuals with intellectual disabilities that we do not cover. Finally, although we reviewed articles that included both US and non-US studies, our review of informed consent and federal laws were limited to those in the United States.

In conclusion, the issues highlighted in our review of the literature provide a framework for researchers to consider when conducting EHR research with individuals with intellectual disabilities. Future research should focus on further understanding the ethical, legal, and social issues in this type of research by asking individuals with intellectual disabilities, and their parents, for direct input on their preferences. Both qualitative and quantitative approaches will provide valuable information for researchers, so they can design studies that are more inclusive of this population using appropriate safeguards and protections, as needed.

Conflicts of Interest

None declared.

References

1. Prokop JW, May T, Strong K, Bilinovich SM, Bupp C, Rajasekaran S, et al. Genome sequencing in the clinic: the past, present, and future of genomic medicine. Physiol Genomics 2018 Aug 1;50(8):563-579 [FREE Full text] [doi: 10.1152/physiolgenomics.00046.2018] [Medline: 29727589]
2. Rabbani B, Tekin M, Mahdieh N. The promise of whole-exome sequencing in medical genetics. J Hum Genet 2014 Jan;59(1):5-15. [doi: 10.1038/jhg.2013.114] [Medline: 24196381]
3. Lionel AC, Costain G, Monfared N, Walker S, Reuter MS, Hosseini SM, et al. Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. Genet Med 2018 Apr;20(4):435-443 [FREE Full text] [doi: 10.1038/gim.2017.119] [Medline: 28771251]
4. Vissers LE, van Nimwegen KJ, Schieving JH, Kamsteeg E, Kleefstra T, Yntema HG, et al. A clinical utility study of exome sequencing versus conventional genetic testing in pediatric neurology. Genet Med 2017 Sep;19(9):1055-1063 [FREE Full text] [doi: 10.1038/gim.2017.1] [Medline: 28333917]
5. Franceschini N, Frick A, Kopp JB. Genetic testing in clinical settings. Am J Kidney Dis 2018 Oct;72(4):569-581 [FREE Full text] [doi: 10.1053/j.ajkd.2018.02.351] [Medline: 29655499]
6. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nat Rev Genet 2012 May 2;13(6):395-405. [doi: 10.1038/nrg3208] [Medline: 22549152]
7. Wysocki T, Diaz MC, Crutchfield JH, Franciosi JP, Werk LN. Electronic health record as a research tool: Frequency of exposure to targeted clinical problems and health care providers' clinical proficiency. J Biomed Inform 2017 Jun;70:14-26 [FREE Full text] [doi: 10.1016/j.jbi.2017.04.016] [Medline: 28442433]
8. Cowie MR, Blomster JI, Curtis LH, Duclaux S, Ford I, Fritz F, et al. Electronic health records to facilitate clinical research. Clin Res Cardiol 2017 Jan;106(1):1-9 [FREE Full text] [doi: 10.1007/s00392-016-1025-6] [Medline: 27557678]
9. Bowton E, Field JR, Wang S, Schildcrout JS, van Driest SL, Delaney JT, et al. Biobanks and electronic medical records: enabling cost-effective research. Sci Transl Med 2014 Apr 30;6(234):234cm3 [FREE Full text] [doi: 10.1126/scitranslmed.3008604] [Medline: 24786321]
10. Gottesman O, Kuivaniemi H, Tromp G, Faucett WA, Li R, Manolio TA, eMERGE Network. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. Genet Med 2013 Oct;15(10):761-771 [FREE Full text] [doi: 10.1038/gim.2013.72] [Medline: 23743551]
11. Kho AN, Pacheco JA, Peissig PL, Rasmussen L, Newton KM, Weston N, et al. Electronic medical records for genetic research: results of the eMERGE consortium. Sci Transl Med 2011 Apr 20;3(79):79re1 [FREE Full text] [doi: 10.1126/scitranslmed.3001807] [Medline: 21508311]
12. McCarty CA, Chisholm RL, Chute CG, Kullo IJ, Jarvik GP, Larson EB, eMERGE Team. The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. BMC Med Genomics 2011 Jan 26;4:13 [FREE Full text] [doi: 10.1186/1755-8794-4-13] [Medline: 21269473]
13. Carey DJ, Fetterolf SN, Davis FD, Faucett WA, Kirchner HL, Mirshahi U, et al. The Geisinger MyCode community health initiative: an electronic health record-linked biobank for precision medicine research. Genet Med 2016 Sep;18(9):906-913 [FREE Full text] [doi: 10.1038/gim.2015.187] [Medline: 26866580]
14. National Institutes of Health (NIH) — All of Us. 2016 Oct 13. NIH Funds Additional Medical Centers to Expand National Precision Medicine Research Program URL: https://allofus.nih.gov/news-events-and-media/announcements/nih-funds-additional-medical-centers-expand-national-precision [accessed 2020-03-17]

15. Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. Bioinformatics 2010 May 1;26(9):1205-1210 [FREE Full text] [doi: 10.1093/bioinformatics/btq126] [Medline: 20335276]
16. Hazin R, Brothers KB, Malin BA, Koenig BA, Sanderson SC, Rothstein MA, et al. Ethical, legal, and social implications of incorporating genomic information into electronic health records. Genet Med 2013 Oct;15(10):810-816 [FREE Full text] [doi: 10.1038/gim.2013.117] [Medline: 24030434]
17. Sabatello M. Precision medicine, health disparities, and ethics: the case for disability inclusion. Genet Med 2018 Apr;20(4):397-399 [FREE Full text] [doi: 10.1038/gim.2017.120] [Medline: 28771252]
18. Spong CY, Bianchi DW. Improving public health requires inclusion of underrepresented populations in research. J Am Med Assoc 2018 Jan 23;319(4):337-338. [doi: 10.1001/jama.2017.19138] [Medline: 29285540]
19. Sankar PL, Parker LS. The Precision Medicine Initiative's All of Us Research Program: an agenda for research on its ethical, legal, and social issues. Genet Med 2017 Jul;19(7):743-750. [doi: 10.1038/gim.2016.183] [Medline: 27929525]
20. Sabatello M, Chen Y, Zhang Y, Appelbaum PS. Disability inclusion in precision medicine research: a first national survey. Genet Med 2019 Oct;21(10):2319-2327. [doi: 10.1038/s41436-019-0486-1] [Medline: 30899094]
21. Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc 2015 Sep;13(3):141-146. [doi: 10.1097/XEB.0000000000000050] [Medline: 26134548]
22. Peterson J, Pearce PF, Ferguson LA, Langford CA. Understanding scoping reviews: Definition, purpose, and process. J Am Assoc Nurse Pract 2017 Jan;29(1):12-16. [doi: 10.1002/2327-6924.12380] [Medline: 27245885]
23. Botkin JR. Informed consent for genetic research. Curr Protoc Hum Genet 2010 Jul;Chapter 1:Unit 1.16. [doi: 10.1002/0471142905.hg0116s66] [Medline: 20582914]
24. Haga SB, Beskow LM. Ethical, legal, and social implications of biobanks for genetics research. Adv Genet 2008;60:505-544. [doi: 10.1016/S0065-2660(07)00418-X] [Medline: 18358331]
25. eCFR — Code of Federal Regulations. Part 46—Protection of Human Subjects URL: https://www.ecfr.gov/cgi-bin/retrieveECFR?gp=&SID=83cd09e1c0f5c6937cd9d7513160fc3f&pitd=20180719&n=pt45.1.46&r=PART&ty=HTML [accessed 2020-03-17]
26. Miller FG. Research on medical records without informed consent. J Law Med Ethics 2008;36(3):560-566. [doi: 10.1111/j.1748-720X.2008.304.x] [Medline: 18840249]
27. Layman EJ. Ethical issues and the electronic health record. Health Care Manag 2008;27(2):165-176. [doi: 10.1097/01.hcm.0000285044.19666.a8]
28. Hoffman S, Podgurski A. Balancing privacy, autonomy, and scientific needs in electronic health records research. SMU Law Rev 2012;65(1):85-144. [Medline: 30028097]
29. Largent EA, Joffe S, Miller FG. Can research and care be ethically integrated? Hastings Cent Rep 2011;41(4):37-46. [doi: 10.1002/j.1552-146x.2011.tb00123.x] [Medline: 21845922]
30. -. ASHG report. Statement on informed consent for genetic research. The American Society of Human Genetics. Am J Hum Genet 1996 Aug;59(2):471-474 [FREE Full text] [Medline: 8755936]
31. National Bioethics Advisory Commission. Bioethics Research Library Archives. 1999 Aug. Research Involving Human Biological Materials: Ethical Issues and Policy Guidance. Volume 1 URL: https://bioethicsarchive.georgetown.edu/nbac/hbm.pdf [accessed 2019-10-11]
32. McGuire AL, Beskow LM. Informed consent in genomics and genetic research. Annu Rev Genomics Hum Genet 2010;11:361-381 [FREE Full text] [doi: 10.1146/annurev-genom-082509-141711] [Medline: 20477535]
33. Wolf LE, Lo B. Untapped potential: IRB guidance for the ethical research use of stored biological materials. IRB 2004;26(4):1-8. [Medline: 15449408]
34. Lunshof JE, Chadwick R, Vorhaus DB, Church GM. From genetic privacy to open consent. Nat Rev Genet 2008 May;9(5):406-411. [doi: 10.1038/nrg2360] [Medline: 18379574]
35. Grizzle W, Grody WW, Noll WW, Sobel ME, Stass SA, Trainer T, et al. Recommended policies for uses of human tissue in research, education, and quality control. Ad Hoc Committee on Stored Tissue, College of American Pathologists. Arch Pathol Lab Med 1999 Apr;123(4):296-300. [doi: 10.1043/0003-9985(1999)123<0296:RPFUOH>2.0.CO;2] [Medline: 10320140]
36. Wendler D. One-time general consent for research on biological samples. Br Med J 2006 Mar 4;332(7540):544-547 [FREE Full text] [doi: 10.1136/bmj.332.7540.544] [Medline: 16513715]
37. Kaye J, Whitley EA, Lund D, Morrison M, Teare H, Melham K. Dynamic consent: a patient interface for twenty-first century research networks. Eur J Hum Genet 2015 Feb;23(2):141-146 [FREE Full text] [doi: 10.1038/ejhg.2014.71] [Medline: 24801761]
38. NIH Office of Science Policy. NIH Genomic Data Sharing URL: https://osp.od.nih.gov/scientific-sharing/genomic-data-sharing/ [accessed 2019-10-11]
39. Greely HT. The uneasy ethical and legal underpinnings of large-scale genomic biobanks. Annu Rev Genomics Hum Genet 2007;8:343-364. [doi: 10.1146/annurev.genom.7.080505.115721] [Medline: 17550341]
40. Hansson MG, Dillner J, Bartram CR, Carlson JA, Helgesson G. Should donors be allowed to give broad consent to future biobank research? Lancet Oncol 2006 Mar;7(3):266-269. [doi: 10.1016/S1470-2045(06)70618-0] [Medline: 16510336]

41. Caulfield T, Murdoch B. Genes, cells, and biobanks: Yes, there's still a consent problem. PLoS Biol 2017 Jul;15(7):e2002654 [FREE Full text] [doi: 10.1371/journal.pbio.2002654] [Medline: 28742850]
42. Willison DJ, Emerson C, Szala-Meneok KV, Gibson E, Schwartz L, Weisbaum KM, et al. Access to medical records for research purposes: varying perceptions across research ethics boards. J Med Ethics 2008 Apr;34(4):308-314. [doi: 10.1136/jme.2006.020032] [Medline: 18375687]
43. Sanderson SC, Brothers KB, Mercaldo ND, Clayton EW, Antommaria AH, Aufox SA, et al. Public attitudes toward consent and data sharing in biobank research: a large multi-site experimental survey in the US. Am J Hum Genet 2017 Mar 2;100(3):414-427 [FREE Full text] [doi: 10.1016/j.ajhg.2017.01.021] [Medline: 28190457]
44. Goodman D, Johnson CO, Wenzel L, Bowen D, Condit C, Edwards KL. Consent issues in genetic research: Views of research participants. Public Health Genomics 2016;19(4):220-228 [FREE Full text] [doi: 10.1159/000447346] [Medline: 27376949]
45. Platt J, Bollinger J, Dvoskin R, Kardia SL, Kaufman D. Public preferences regarding informed consent models for participation in population-based genomic research. Genet Med 2014 Jan;16(1):11-18 [FREE Full text] [doi: 10.1038/gim.2013.59] [Medline: 23660530]
46. McGuire AL, Oliver JM, Slashinski MJ, Graves JL, Wang T, Kelly PA, et al. To share or not to share: a randomized trial of consent for data sharing in genome research. Genet Med 2011 Nov;13(11):948-955 [FREE Full text] [doi: 10.1097/GIM.0b013e3182227589] [Medline: 21785360]
47. Robinson JO, Slashinski MJ, Wang T, Hilsenbeck SG, McGuire AL. Participants' recall and understanding of genomic research and large-scale data sharing. J Empir Res Hum Res Ethics 2013 Oct;8(4):42-52 [FREE Full text] [doi: 10.1525/jer.2013.8.4.42] [Medline: 24169421]
48. Campbell B, Thomson H, Slater J, Coward C, Wyatt K, Sweeney K. Extracting information from hospital records: what patients think about consent. Qual Saf Health Care 2007 Dec;16(6):404-408 [FREE Full text] [doi: 10.1136/qshc.2006.020313] [Medline: 18055882]
49. Hoop JG, Roberts LW, Hammond KA. Genetic testing of stored biological samples: views of 570 US workers. Genet Test Mol Biomarkers 2009 Jun;13(3):331-337. [doi: 10.1089/gtmb.2008.0117] [Medline: 19405873]
50. Buckley BS, Murphy AW, MacFarlane AE. Public attitudes to the use in research of personal health information from general practitioners' records: a survey of the Irish general public. J Med Ethics 2011 Jan;37(1):50-55. [doi: 10.1136/jme.2010.037903] [Medline: 21071570]
51. Clerkin P, Buckley BS, Murphy AW, MacFarlane AE. Patients' views about the use of their personal information from general practice medical records in health research: a qualitative study in Ireland. Fam Pract 2013 Feb;30(1):105-112. [doi: 10.1093/fampra/cms036] [Medline: 22850249]
52. Hill EM, Turner EL, Martin RM, Donovan JL. 'Let's get the best quality research we can': public awareness and acceptance of consent to use existing data in health research: a systematic review and qualitative study. BMC Med Res Methodol 2013 Jun 4;13:72 [FREE Full text] [doi: 10.1186/1471-2288-13-72] [Medline: 23734773]
53. Toccaceli V, Fagnani C, Stazi MA. Medical records confidentiality and public health research: two values at stake? An italian survey focus on individual preferences. J Public Health Res 2015 Feb 20;4(1):401 [FREE Full text] [doi: 10.4081/jphr.2015.401] [Medline: 25918693]
54. Willison D, Keshavjee K, Nair K, Goldsmith C, Holbrook A, Computerization of Medical Practices for the Enhancement of Therapeutic Effectiveness investigators. Patients' consent preferences for research uses of information in electronic medical records: interview and survey data. Br Med J 2003 Feb 15;326(7385):373 [FREE Full text] [doi: 10.1136/bmj.326.7385.373] [Medline: 12586673]
55. Kluge EW. Security and privacy of EHR systems--ethical, social and legal requirements. Stud Health Technol Inform 2003;96:121-127. [Medline: 15061533]
56. Damschroder LJ, Pritts JL, Neblo MA, Kalarickal RJ, Creswell JW, Hayward RA. Patients, privacy and trust: patients' willingness to allow researchers to access their medical records. Soc Sci Med 2007 Jan;64(1):223-235. [doi: 10.1016/j.socscimed.2006.08.045] [Medline: 17045717]
57. Robling MR, Hood K, Houston H, Pill R, Fay J, Evans HM. Public attitudes towards the use of primary care patient record data in medical research without consent: a qualitative study. J Med Ethics 2004 Feb;30(1):104-109 [FREE Full text] [doi: 10.1136/jme.2003.005157] [Medline: 14872086]
58. Trinidad SB, Fullerton SM, Bares JM, Jarvik GP, Larson EB, Burke W. Genomic research and wide data sharing: views of prospective participants. Genet Med 2010 Aug;12(8):486-495 [FREE Full text] [doi: 10.1097/GIM.0b013e3181e38f9e] [Medline: 20535021]
59. Perera G, Holbrook A, Thabane L, Foster G, Willison DJ. Views on health information sharing and privacy from primary care practices using electronic medical records. Int J Med Inform 2011 Feb;80(2):94-101. [doi: 10.1016/j.ijmedinf.2010.11.005] [Medline: 21167771]
60. Caine K, Hanania R. Patients want granular privacy control over health information in electronic medical records. J Am Med Inform Assoc 2013 Jan 1;20(1):7-15 [FREE Full text] [doi: 10.1136/amiajnl-2012-001023] [Medline: 23184192]

61. Tierney WM, Alpert SA, Byrket A, Caine K, Leventhal JC, Meslin EM, et al. Provider responses to patients controlling access to their electronic health records: a prospective cohort study in primary care. J Gen Intern Med 2015 Jan;30(Suppl 1):S31-S37 [FREE Full text] [doi: 10.1007/s11606-014-3053-0] [Medline: 25480720]
62. Hull SC, Sharp RR, Botkin JR, Brown M, Hughes M, Sugarman J, et al. Patients' views on identifiability of samples and informed consent for genetic research. Am J Bioeth 2008 Oct;8(10):62-70 [FREE Full text] [doi: 10.1080/15265160802478404] [Medline: 19003716]
63. Badzek L, Henaghan M, Turner M, Monsen R. Ethical, legal, and social issues in the translation of genomics into health care. J Nurs Scholarsh 2013 Mar;45(1):15-24. [doi: 10.1111/jnu.12000] [Medline: 23369261]
64. Chow-White PA, MacAulay M, Charters A, Chow P. From the bench to the bedside in the big data age: ethics and practices of consent and privacy for clinical genomics and personalized medicine. Ethics Inf Technol 2015 Aug;17(3):189-200. [doi: 10.1007/s10676-015-9373-x]
65. Tavani HT. Genomic research and data-mining technology: implications for personal privacy and informed consent. Ethics Inf Technol 2004;6(1):15-28. [doi: 10.1023/b:etin.0000036156.77169.31] [Medline: 16969958]
66. El Emam K, Jabbouri S, Sams S, Drouet Y, Power M. Evaluating common de-identification heuristics for personal health information. J Med Internet Res 2006 Nov 21;8(4):e28 [FREE Full text] [doi: 10.2196/jmir.8.4.e28] [Medline: 17213047]
67. Malin B, Loukides G, Benitez K, Clayton EW. Identifiability in biobanks: models, measures, and mitigation strategies. Hum Genet 2011 Sep;130(3):383-392 [FREE Full text] [doi: 10.1007/s00439-011-1042-5] [Medline: 21739176]
68. Kimball BC, Nowakowski KE, Maschke KJ, McCormick JB. Genomic data in the electronic medical record: perspectives from a biobank community advisory board. J Empir Res Hum Res Ethics 2014 Dec;9(5):16-24. [doi: 10.1177/1556264614553922] [Medline: 25747687]
69. Cuticchia AJ. Applying existing ethical principles to personal medicine. Internet J Law Healthc Ethics 2009;6(1):3-31. [doi: 10.5580/18d1]
70. National Heart, Lung, and Blood Institute working group, Fabsitz RR, McGuire A, Sharp RR, Puggal M, Beskow LM, et al. Ethical and practical guidelines for reporting genetic research results to study participants: updated guidelines from a National Heart, Lung, and Blood Institute working group. Circ Cardiovasc Genet 2010 Dec;3(6):574-580 [FREE Full text] [doi: 10.1161/CIRCGENETICS.110.958827] [Medline: 21156933]
71. Secretary's Advisory Committee on Human Research Protections (SACHRP). The US Department of Health and Human Services (HHS). Attachment B: Return of Individual Research Results URL: https://www.hhs.gov/ohrp/sachrp-committee/recommendations/attachment-b-return-individual-research-results/index.html [accessed 2019-11-10]
72. Fullerton SM, Wolf WA, Brothers KB, Clayton EW, Crawford DC, Denny JC, et al. Return of individual research results from genome-wide association studies: experience of the Electronic Medical Records and Genomics (eMERGE) Network. Genet Med 2012 Apr;14(4):424-431 [FREE Full text] [doi: 10.1038/gim.2012.15] [Medline: 22361898]
73. Pereira S, Robinson JO, McGuire AL. Return of individual genomic research results: what do consent forms tell participants? Eur J Hum Genet 2016 Nov;24(11):1524-1529 [FREE Full text] [doi: 10.1038/ejhg.2016.76] [Medline: 27329739]
74. Purdy S, Finkelstein JA, Fletcher R, Christiansen C, Inui TS. Patient participation in research in the managed care environment: key perceptions of members in an HMO. J Gen Intern Med 2000 Jul;15(7):492-495 [FREE Full text] [doi: 10.1046/j.1525-1497.2000.07025.x] [Medline: 10940136]
75. Ruiz-Canela M, Valle-Mansilla JI, Sulmasy DP. What research participants want to know about genetic research results: the impact of 'genetic exceptionalism'. J Empir Res Hum Res Ethics 2011 Sep;6(3):39-46. [doi: 10.1525/jer.2011.6.3.39] [Medline: 21931236]
76. Avard DM, Knoppers BM. Ethical dimensions of genetics in pediatric neurology: a look into the future. Semin Pediatr Neurol 2002 Mar;9(1):53-61. [doi: 10.1053/spen.2002.30342] [Medline: 11931127]
77. Coors ME, Raymond KM. Substance use disorder genetic research: investigators and participants grapple with the ethical issues. Psychiatr Genet 2009 Apr;19(2):83-90 [FREE Full text] [doi: 10.1097/YPG.0b013e328320800e] [Medline: 19668113]
78. Goldenberg AJ, Hull SC, Botkin JR, Wilfond BS. Pediatric biobanks: approaching informed consent for continuing research after children grow up. J Pediatr 2009 Oct;155(4):578-583 [FREE Full text] [doi: 10.1016/j.jpeds.2009.04.034] [Medline: 19595370]
79. Ries NM, LeGrandeur J, Caulfield T. Handling ethical, legal and social issues in birth cohort studies involving genetic research: responses from studies in six countries. BMC Med Ethics 2010 Mar 23;11:4 [FREE Full text] [doi: 10.1186/1472-6939-11-4] [Medline: 20331891]
80. Brothers KB, Clayton EW. Parental perspectives on a pediatric human non-subjects biobank. AJOB Prim Res 2012 Jan 1;3(3):21-29 [FREE Full text] [doi: 10.1080/21507716.2012.662576] [Medline: 23181193]
81. Doshi-Velez F, Ge Y, Kohane I. Comorbidity clusters in autism spectrum disorders: an electronic health record time-series analysis. Pediatrics 2014 Jan;133(1):e54-e63 [FREE Full text] [doi: 10.1542/peds.2013-0819] [Medline: 24323995]
82. Angell AM, Solomon O. The social life of health records: understanding families' experiences of autism. Soc Sci Med 2014 Sep;117:50-57 [FREE Full text] [doi: 10.1016/j.socscimed.2014.07.020] [Medline: 25042544]
83. Croen LA, Zerbo O, Qian Y, Massolo ML, Rich S, Sidney S, et al. The health status of adults on the autism spectrum. Autism 2015 Oct;19(7):814-823. [doi: 10.1177/1362361315577517] [Medline: 25911091]

84. Simon GE, Unützer J, Young BE, Pincus HA. Large medical databases, population-based research, and patient confidentiality. Am J Psychiatry 2000 Nov;157(11):1731-1737. [doi: 10.1176/appi.ajp.157.11.1731] [Medline: 11058465]
85. Carpenter WJ, Gold J, Lahti A, Queern CA, Conley RR, Bartko JJ, et al. Decisional capacity for informed consent in schizophrenia research. Arch Gen Psychiatry 2000 Jun;57(6):533-538. [doi: 10.1001/archpsyc.57.6.533] [Medline: 10839330]
86. Dunn LB, Nowrangi MA, Palmer BW, Jeste DV, Saks ER. Assessing decisional capacity for clinical research or treatment: a review of instruments. Am J Psychiatry 2006 Aug;163(8):1323-1334. [doi: 10.1176/ajp.2006.163.8.1323] [Medline: 16877642]
87. Jefferson A, Lambe S, Moser D, Byerly L, Ozonoff A, Karlawish J. Decisional capacity for research participation in individuals with mild cognitive impairment. J Am Geriatr Soc 2008 Jul;56(7):1236-1243 [FREE Full text] [doi: 10.1111/j.1532-5415.2008.01752.x] [Medline: 18482298]
88. Moye J, Karel MJ, Gurrera RJ, Azar AR. Neuropsychological predictors of decision-making capacity over 9 months in mild-to-moderate dementia. J Gen Intern Med 2006 Jan;21(1):78-83 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.00288.x] [Medline: 16423129]
89. Iacono T, Murray V. Issues of informed consent in conducting medical research involving people with intellectual disability. J Appl Res Int Dis 2003;16(1):41-51. [doi: 10.1046/j.1468-3148.2003.00141.x]
90. McDonald KE, Kidney CA. What is right? Ethics in intellectual disabilities research. J Policy Pract Intellect Disabil 2012 Mar;9(1):27-39. [doi: 10.1111/j.1741-1130.2011.00319.x]
91. Wheeler AC, Wylie A, Raspa M, Villagomez A, Miller K, Edwards A, et al. Decisional capacity for informed consent in males and females with fragile X syndrome. J Autism Dev Disord 2019 Mar 1:- [Epub ahead of print]. [doi: 10.1007/s10803-019-03930-4] [Medline: 30825084]
92. Fields LM, Calvert JD. Informed consent procedures with cognitively impaired patients: A review of ethics and best practices. Psychiatry Clin Neurosci 2015 Aug;69(8):462-471 [FREE Full text] [doi: 10.1111/pcn.12289] [Medline: 25756740]
93. Golnik A, Maccabee-Ryaboy N, Scal P, Wey A, Gaillard P. Shared decision making: improving care for children with autism. Intellect Dev Disabil 2012 Aug;50(4):322-331. [doi: 10.1352/1934-9556-50.4.322] [Medline: 22861133]
94. Wullink M, Widdershoven G, Valk HS, Metsemakers J, Dinant G. Autonomy in relation to health among people with intellectual disability: a literature review. J Intellect Disabil Res 2009 Sep;53(9):816-826. [doi: 10.1111/j.1365-2788.2009.01196.x] [Medline: 19646099]
95. Griggs RC, Batshaw M, Dunkle M, Gopal-Srivastava R, Kaye E, Krischer J, Rare Diseases Clinical Research Network. Clinical research for rare disease: opportunities, challenges, and solutions. Mol Genet Metab 2009 Jan;96(1):20-26 [FREE Full text] [doi: 10.1016/j.ymgme.2008.10.003] [Medline: 19013090]
96. McCormack P, Kole A, Gainotti S, Mascalzoni D, Molster C, Lochmüller H, et al. 'You should at least ask'. The expectations, hopes and fears of rare disease patients on large-scale data and biomaterial sharing for genomics research. Eur J Hum Genet 2016 Oct;24(10):1403-1408 [FREE Full text] [doi: 10.1038/ejhg.2016.30] [Medline: 27049302]
97. Tambuyzer E. Rare diseases, orphan drugs and their regulation: questions and misconceptions. Nat Rev Drug Discov 2010 Dec;9(12):921-929. [doi: 10.1038/nrd3275] [Medline: 21060315]
98. Wästfelt M, Fadeel B, Henter JI. A journey of hope: lessons learned from studies on rare diseases and orphan drugs. J Intern Med 2006 Jul;260(1):1-10 [FREE Full text] [doi: 10.1111/j.1365-2796.2006.01666.x] [Medline: 16789973]
99. Hansson MG, Lochmüller H, Riess O, Schaefer F, Orth M, Rubinstein Y, et al. The risk of re-identification versus the need to identify individuals in rare disease research. Eur J Hum Genet 2016 Nov;24(11):1553-1558 [FREE Full text] [doi: 10.1038/ejhg.2016.52] [Medline: 27222291]
100. Gilbert T. Involving people with learning disabilities in research: issues and possibilities. Health Soc Care Community 2004 Jul;12(4):298-308. [doi: 10.1111/j.1365-2524.2004.00499.x] [Medline: 15272885]
101. Chen A. The Verge. 2019 Jun 29. Why It's Time to Rethink the Laws That Keep Our Health Data Private URL: https://www.theverge.com/2019/1/29/18197541/health-data-privacy-hipaa-policy-business-science [accessed 2019-11-10]
102. Haga SB, Mills R, Pollak KI, Rehder C, Buchanan AH, Lipkus IM, et al. Developing patient-friendly genetic and genomic test reports: formats to promote patient engagement and understanding. Genome Med 2014;6(7):58 [FREE Full text] [doi: 10.1186/s13073-014-0058-6] [Medline: 25473429]

Abbreviations

EHR: electronic health record
eMERGE: Electronic Medical Records and Genomics
HIPAA: Health Insurance Portability and Accountability Act
HMO: Health Maintenance Organization
IRB: institutional review board

Edited by G Eysenbach; submitted 18.10.19; peer-reviewed by S Veeranki, D Wong; comments to author 10.12.19; revised version received 17.01.20; accepted 22.02.20; published 21.05.20.

Please cite as:
Raspa M, Moultrie R, Wagner L, Edwards A, Andrews S, Frisch MK, Turner-Brown L, Wheeler A
Ethical, Legal, and Social Issues Related to the Inclusion of Individuals With Intellectual Disabilities in Electronic Health Record Research: Scoping Review
J Med Internet Res 2020;22(5):e16734
URL: http://www.jmir.org/2020/5/e16734/
doi:10.2196/16734
PMID:32436848

Review

The Effectiveness of Electronic Health Interventions for Promoting HIV-Preventive Behaviors Among Men Who Have Sex With Men: Meta-Analysis Based on an Integrative Framework of Design and Implementation Features

Meiqi Xin[1], PhD; Kasisomayajula Viswanath[2], PhD; Angela Yuen-Chun Li[1], BSocSc; Wangnan CAO[3], PhD; Yuhong HU[4], BMed; Joseph Tak-Fai Lau[1], FFPH, PhD; Phoenix Kit-Han Mo[1], MSc, PhD, CPsychol

[1]Jockey Club School of Public Health and Primary Care, The Chinese University of Hong Kong, Hong Kong, China (Hong Kong)

[2]TH Chan School of Public Health, Harvard University, Boston, MA, United States

[3]Center for Evidence Synthesis in Health, School of Public Health, Brown University, Providence, RI, United States

[4]School of Public Health, Shanghai University of Traditional Chinese Medicine, Shanghai, China

Corresponding Author:
Phoenix Kit-Han Mo, MSc, PhD, CPsychol
Jockey Club School of Public Health and Primary Care
The Chinese University of Hong Kong
Ngan Shing St, Shatin
Hong Kong
China (Hong Kong)
Phone: 852 22528765
Email: phoenix.mo@cuhk.edu.hk

Abstract

Background: The disproportionately high prevalence of HIV among men who have sex with men (MSM) is a global concern. Despite the increasing utilization of electronic health (eHealth) technology in the delivery of HIV prevention interventions, few studies have systematically explored its effectiveness and association with various intervention characteristics.

Objective: This study aimed to conduct a meta-analysis of the effectiveness of eHealth technology–based interventions for promoting HIV-preventive behaviors among MSM and to determine effectiveness predictors within a framework integrating design and implementation features.

Methods: A systematic literature search using terms related to eHealth technology, HIV, the MSM population, and an experimental study design was performed using 5 databases (ie, MEDLINE, PsycINFO, EMBASE, Web of Science, and ProQuest Dissertations & Theses) and other sources (eg, bibliographies of relevant reviews and JMIR Publications). First, primary meta-analyses were conducted to estimate the effectiveness of eHealth interventions (d_+) in changing 3 HIV-preventive behaviors among MSM: unprotected anal intercourse (UAI), HIV testing, and multiple sex partnership (MSP). Moderation analyses were then conducted to examine a priori effectiveness predictors including behavioral treatment components (eg, theory use, tailoring strategy use, navigation style, and treatment duration), eHealth technology components (eg, operation mode and modality type), and intervention adherence.

Results: A total of 46 studies were included. The overall effect sizes at end point were small but significant for all outcomes (UAI: d_+=−.21, P<.001; HIV testing: d_+=.38, P<.001; MSP: d_+=−.26, P=.02). The intervention effects on UAI were significantly larger when compared with preintervention groups than with concurrent groups. Greater UAI reductions were associated with the increased use of tailoring strategies, provision of feedback, and tunneling navigation in interventions with a concurrent group, whereas reductions were associated with the use of self-paced navigation in interventions with a preintervention group. Greater uptake of HIV testing was associated with longer treatment duration; computer-mediated communication; and the use of messaging, social media, or a combined technology modality. Higher intervention adherence consistently predicted larger effects on UAI and HIV testing.

Conclusions: This study provided empirical evidence for the effectiveness of eHealth interventions in promoting HIV-preventive behaviors among MSM. Features of treatment content and eHealth technology might best predict the intervention effects on UAI

and HIV testing, respectively. Most importantly, intervention adherence tended to play an important role in achieving better effectiveness. The findings could help inform the development of efficacious interventions for HIV prevention in the future.

(J Med Internet Res 2020;22(5):e15977) doi:10.2196/15977

KEYWORDS

HIV; AIDS; sexual and gender minorities; telemedicine; meta-analysis; systematic review

Introduction

Background

In marked contrast to a declining trend in its global burden in the era of potent antiretroviral therapy, the HIV epidemic has continued to expand in the men who have sex with men (MSM) population across countries of all incomes in recent years [1,2]. The disproportionate burden is evidenced to be driven by a stable increase in HIV risk behaviors [3,4], which may offset the benefits of improved treatment coverage and biomedical advances [3,5]. Undiagnosed infections or infections diagnosed late arising from the relatively low uptake of HIV testing also fuel the epidemic by hampering treatment delivery and increasing transmission [6,7]. Therefore, the promotion of HIV-preventive behaviors is an indispensable part of comprehensive prevention efforts to control HIV [8]. Behavioral interventions among MSM have been found to significantly reduce risky sexual behaviors and increase HIV testing [9,10].

The global response to the ongoing epidemic warrants the application of innovative technology to develop efficacious behavioral interventions [7]. Electronic health (eHealth) is well recognized as the use of information and communication technologies for health, including websites, computerized programs or apps, social networking sites or chatrooms, email, or text messaging, that feature internet connectivity or digital interactivity via computer or mobile devices [11-13]. Previous quantitative reviews have shown that the impact of eHealth technology–based interventions on HIV prevention–related behaviors and theoretical correlates was significant but varied as a function of intervention characteristics [14,15].

New technology has been extensively used by MSM to socialize and seek sexual partners and to access information on sexual health [16,17]. An early meta-analysis revealed a significant effect of sexual health interventions delivered via interactive digital media for the overall population but not MSM, although a growing body of evidence has been accumulated since then [12]. Despite recent synthesis attempts focusing on MSM [18,19], the evaluation of effect magnitudes across outcomes and the disentanglement of efficacious components from complex intervention designs have been limited by the qualitative nature of the reviews. There is a lack of meta-analytic reviews that would allow rigorous testing of the a priori factors that predict intervention effectiveness [20] and thus inform the design and implementation of future programs.

Potential Predictors of Electronic Health Intervention Effectiveness

An eHealth technology–based behavioral intervention includes 2 elements: *behavioral treatment* (what type of intervention is designed and programmed to target behavior change) and *eHealth technology* (how the treatment is delivered via eHealth platforms). This study proposed a conceptualized framework to capture critical factors that are theoretically and empirically demonstrated to explain the effectiveness of eHealth interventions.

Behavioral Treatment Components

Treatment content is characterized by the *use of theory* for intervention development, as theoretical constructs can be operationalized into treatment techniques [21]. Significantly greater reductions in unprotected anal intercourse (UAI) among MSM have been observed for interventions reporting any use of theory [9], yet it is unclear whether the extent of use predicts intervention effectiveness, particularly in the eHealth context. A meta-analysis of internet-based behavioral interventions across health domains revealed a significant association between more extensive theory use and increased effect sizes [22].

Treatment content also features the application of *tailoring strategies*, defined as the integration of recipients' responses into the intervention system to generate user-driven content [23]. There are 3 tailoring types: (1) *feedback*: providing unique recommendations derived from an assessment of individual needs or characteristics related to a given behavior; (2) *adaptation*: matching content to a relevant group based on known behavioral determinants; and (3) *personalization*: customizing content with personally identifiable information [24]. Computer technology–based interventions with individually tailored content adapted to the stage of change were shown to be more efficacious in increasing condom use in the general population [14].

The programming features of a treatment can strongly determine the intervention intensity. The amount of accessible content per interaction with the intervention differs by *navigation style*: *tunneled* interventions deliver treatment through a predetermined sequence of steps, whereas *self-paced* interventions release content all at once and permit recipients to control the navigation [25-27]. Existing evidence is inconclusive regarding the influence of navigation style on the effectiveness of eHealth interventions [26,28,29]. *Treatment duration*, referring to the time span of delivery, represents the overall intervention burden. Larger effects and higher adherence have been generally reported for online behavioral interventions of a shorter duration [23].

Electronic Health Technology Components

On the basis of the nature of users' interaction with eHealth platforms, 2 *operation modes* of technology use for intervention delivery have been distinguished: (1) *human-computer interaction (HCI)*, featuring direct interaction with a computerized system and automated delivery of a

preprogrammed treatment and (2) *computer-mediated communication (CMC)*, featuring remote delivery through interpersonal communications via eHealth media. These are considered to be inseparable, albeit essentially different aspects of eHealth technology and capable of deploying similar treatment strategies [25,30]. It has been reported that the provision of remote therapeutic support, not fully automated treatment, achieved significantly better psychological outcomes than passive control [31].

Four *modality types* have been applied to eHealth interventions for HIV prevention, and each of them possesses a unique capacity to facilitate intervention delivery [11,19,32]. The 2 most common modalities are the *interactive module*, in which users actively engage in an intervention following a preset workflow, and the *static site*, in which users passively receive prescriptive information [30]. A previous review suggested the superiority of interactive over static interventions in predicting increased condom use and sexually transmitted infection (STI) testing [12]. The strength of *text messaging* lies in the ubiquitous use of mobile devices to deliver real-time personalized interventions [11,33]. It can incorporate effective communication techniques and has demonstrated its efficacy in promoting health behavior change [33,34]. This study uses the broader term, "messaging," to encompass related technologies that enable multimedia delivery [33]. *Social media* has emerged as a novel modality that features the creation and sharing of user-generated content in an online community [32,35], which can effectively promote HIV testing [35].

Exposure to Intervention Components

Intervention effectiveness observed in real-world practice can change with actual exposure to efficacious intervention components, and nonusage attrition is common in eHealth interventions [36]. *Intervention adherence* is conceptualized as the proportion of participants who engage in the intervention as prescribed to achieve a desired effect [29]. This allows for a comparable measurement of intervention exposure across use parameters, by contrasting the *actual usage* of an intervention during its implementation with the *intended usage* predefined at the design stage [25]. Adherence has been found to significantly predict the effectiveness of eHealth interventions [23].

This study aimed to conduct a meta-analysis of the effectiveness of eHealth technology–based interventions for promoting HIV-preventive behaviors among MSM and to provide an in-depth investigation of effectiveness predictors.

Methods

The guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses were followed [37].

Search and Selection

A systematic search was first performed in the following databases: MEDLINE, PsycINFO, EMBASE, Web of Science, and ProQuest Dissertations & Theses. The search strategy comprised 4 categories of terms related to eHealth technology, HIV or AIDS, the MSM population, and experimental study design, which were tailor-made for each database with restrictions to the English language and human studies (Multimedia Appendix 1). We manually searched bibliographies of relevant reviews, initially retrieved articles, and JMIR Publications to identify additional eligible studies. The eligibility criteria are listed in Table 1.

After removing the duplicates, 3657 relevant articles were identified from all the search sources. The title- and abstract-based eligibility screening was first conducted by 2 independent reviewers, which led to a selection of 143 articles for further full text–based assessment. Finally, 45 articles with a total sample size of 27,704 were deemed eligible and included in this meta-analysis [38-82]. The selection process is shown in Figure 1.

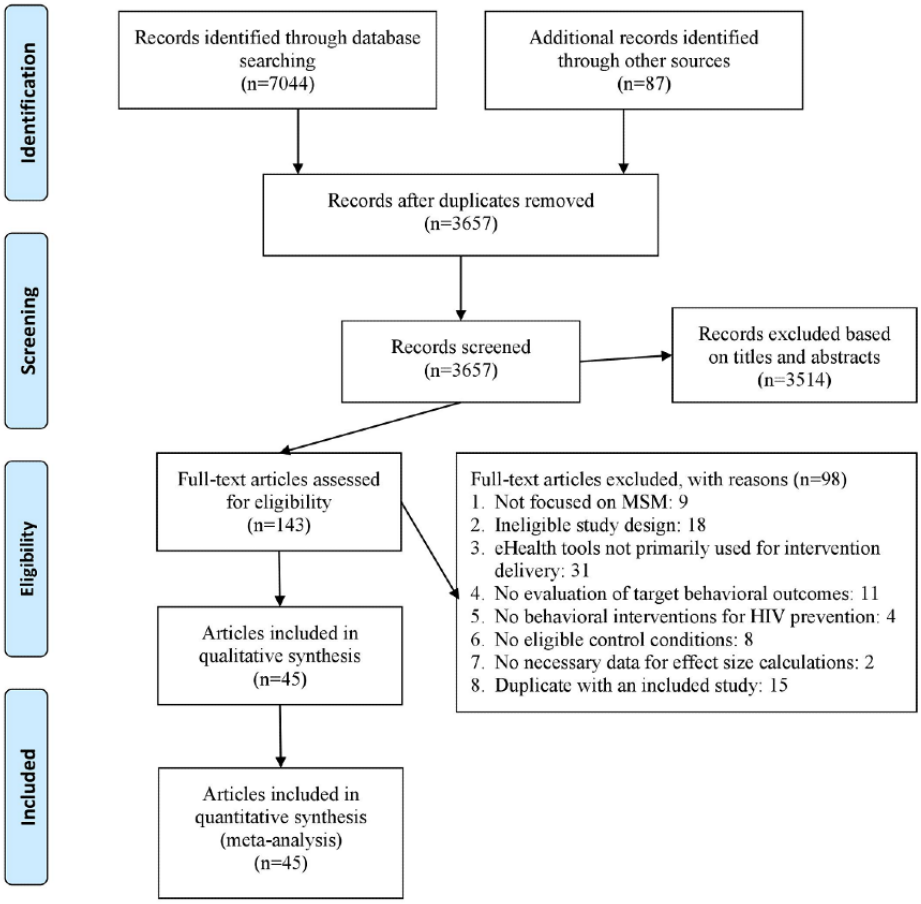
Table 1. Eligibility criteria for elements of a comprehensive search strategy.

Element	Inclusion and exclusion criteria
Population	• Included: being exclusively or primarily (accounting for at least 80% of the sample) focused on MSM[a] or specifically conducting the efficacy evaluation among the MSM subgroup
Intervention	• Included: delivering interventions largely via eHealth[b] technologies, including internet-based tools or interactive computerized programs and administering behavior change interventions for the prevention of HIV infection • Excluded: solely using non-eHealth technologies (eg, telephone-based communications); not specifying the role of eHealth components in a multimedia program; or solely adopting biomedical strategies (eg, pre-exposure prophylaxis) or targeting the HIV care continuum after diagnosis (eg, treatment as prevention) or other sexual health topics (eg, contraception)
Comparator	• Included: using control conditions that are different from interventions in the eHealth components • Excluded: solely aimed at testing a group difference in other factors (eg, sample characteristics or treatment approaches)
Outcome	• Included: measuring HIV-preventive behaviors as efficacy outcomes and providing necessary statistics to estimate the effect size • Excluded: treating HIV-preventive behaviors as confounders or compensatory outcomes (eg, condom use measured in interventions on promoting monogamy); solely measuring behavioral correlates (eg, intentions) or biomarkers of engagement in HIV risk behaviors (eg, sexually transmitted infection occurrence); or computing a composite score for a group of behaviors
Study design	• Included: a randomized controlled trial or nonrandomized experiment • Excluded: a single-group posttest-only experiment; an observational study; or a review or commentary

[a]MSM: men who have sex with men.

[b]eHealth: electronic health.

Figure 1. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart of the screening process. eHealth: electronic health; MSM: men who have sex with men.



Quality Assessment

In all, 44 intervention programs were involved (2 articles reported outcome measures at different follow-up points for the same program). Program quality was evaluated using the assessment tool of Schnall et al [19], which was developed based on the reporting standards for HIV intervention studies established by the HIV Prevention Research Synthesis Team of the US Centers for Disease Control and Prevention [83] and has been well applied in previous reviews [9,19]. An overall percentage score (0%-100%) was computed as the extent to which each program fulfilled the criteria across 7 categories: representativeness, bias and confounding, description of intervention, outcomes and follow-up, statistical analysis, strength of evidence, and group equivalence.

Data Extraction

Program Design

Data were extracted from each eligible program and its supplementary materials by one coder (MX) and double-checked by a second coder (AL). Information was first gathered on factors related to the program design, which were used to define the coding unit and to determine the grouping sets for the primary meta-analysis.

One intervention condition and one comparator were selected to form an independent study (a pairwise comparison) for the meta-analysis. If more than one intervention group was reported, each of them was separately included, given the potential variation in interventions, while the initial control group was proportionally divided up among intervention conditions to ensure independence [84]. If more than one control group was reported, the less intensive one was included for ease of interpretation [85]. If an eligible concurrent comparator was not obtained, or independent group contrasts were not the program focus, the preintervention status was treated as the comparator [86]. Finally, for each established comparison, identification items (eg, author) and sample characteristics (eg, age) were coded.

Study design was classified according to the outcome assessment points and comparison status (ie, using a concurrent or preintervention group). *Comparator type* was coded as (1) passive, including blank, waitlist, and attention controls or (2) active, including non-eHealth and lower tech eHealth treatment. *Follow-up period* was classified according to the interval between the intervention end and the follow-up points: (1) immediately postintervention; (2) short-term: ≤3 months; or (3) long-term follow-up: >3 months. The *outcome measures* with adequate studies for effect size pooling were UAI, condom use, HIV testing, and multiple sex partnership (MSP).

Intervention Features

A cluster of prespecified factors that captured the intervention characteristics were coded within the conceptualized framework described earlier. A relative coding rule was applied, that is, where a certain item was identical across the paired conditions, the study was coded as not having that feature, as it could not explain the conditional difference in the effectiveness.

With regard to the key features of treatment content, the extent of *theory use* was first assessed using the modified Theory Coding Scheme [21]. A total of 11 items pertinent to intervention development were used to compute an overall score [22], with a higher value indicating more extensive use of the theory. The second content feature was the *use of tailoring strategies*, which was coded in both ways, binary (whether adopting the *feedback*, *adaptation*, or *personalization* strategy) and continuous (the number of strategies adopted). The programming-specific treatment features were categorized into mutually exclusive groups: *navigation style* (self-paced or tunneled) and *treatment duration* (single session, ≤1 month, >1-3 months, and >3 months).

The utilization of eHealth technology was characterized by the *operation mode* (ie, HCI and CMC) and the *modality type* (ie, static site, interactive module, messaging, and social media). For both variables, interventions with the presence of more than one feature were grouped into a combined category to allow for exclusive coding. As indicators of intervention exposure, *intended usage* was first extracted and varying levels of *actual usage* were identified accordingly. *Intervention adherence* was then calculated as the percentage of participants whose actual usage matched the intended usage [25].

Data Synthesis

Effect Size Calculation

Standardized mean differences between conditions (Cohen *d*) and their standard error at each follow-up point were derived to represent the magnitude of intervention effects [87,88]. Following Morris and DeShon's [89] procedures, the effect size estimate and its sampling error were computed or transformed to be scaled on a common "raw-score" metric, creating a synthesis across the study design. Unadjusted outcome measurements were retrieved to establish comparability across estimates. For studies in which means and standard deviations were not provided, effect measures reported in other forms (eg, risk ratio) were converted to the Cohen *d* statistic using well-developed calculators [90,91].

As recommended, UAI and condom use were combined into one outcome type, "UAI" [14]. For studies that used multiple instruments to measure an outcome, the most common instrument was chosen (eg, the frequency of UAI was prioritized over the count of UAI partners [56]). For studies measuring different subtypes of a certain outcome on the same scale (eg, serostatus-specific UAI [71]) or reporting subgroup effects by significant moderators that were unrelated to intervention components (eg, affect level [44]), all effect sizes were aggregated meta-analytically within the study. Consequently, each study yielded only one Cohen *d* value per outcome.

Meta-Analysis

With an assumption of the intervention diversity, pooled effect sizes (d_+) and 95% confidence intervals were generated using random effects models with the inverse variance weighting method. Values such as 0.2, 0.5, and 0.8 were interpreted as small, moderate, and large, respectively [92]. Primary meta-analyses were performed to combine effect sizes within different study sets for each outcome: (1) an overall estimate at

the end point (one effect arising per study); (2) group-specific estimates of the end point effect by comparison status (nested groupings); and (3) group-specific estimates by follow-up period (a multi-wave study contributing multiple effects). To maximize analytical power, the first type of estimation based on a full data set was used for further analyses.

Heterogeneity across studies was assessed by using the Q statistic with a *P* value <.05, indicating the presence of significant heterogeneity, and quantified by using the I^2 statistic, where a value of 30% to 60% denoted "moderate" and ≥75% denoted "considerable" heterogeneity [93]. Wherever considerable heterogeneity was found, outliers were identified as studies that significantly distorted the pooled effect using influence analyses [94] and were removed to ensure the accuracy and generalizability of the findings. Publication bias was then assessed by visually inspecting the funnel plot of effect sizes and conducting an Egger regression test to examine the plot asymmetry when there were at least 10 estimates [95].

Next, secondary analyses were performed to test the moderation effect of intervention features where adequate studies were available (n≥10) [96]. Subgroup and meta-regression analyses were applied for dichotomous and continuous measures, respectively; a significant moderator was indicated by a *P* value <.05 for the heterogeneity across subgroups (Q_b) or the regression coefficient (β). All analyses were conducted in R 3.5.2 with the metafor package (Wolfgang Viechtbauer) [97].

Results

Descriptions of Program Characteristics

Over half of the eligible programs (23/44, 52%) were conducted in the United States, and 10 and 8 programs were conducted in Asia and Europe, respectively. Among programs reporting the respective background characteristics, most of the samples had a mean age below 30 years (24/44, 54%), were multiracial or Asian (29/39, 74%), were mainly (≥70%) composed of homosexual males or gays (22/28, 78%), and were non-HIV positive or of mixed (both positive and negative) status (33/36, 92%). In addition, the median overall quality score was 84.2%; the assessment results are provided in detail in Multimedia Appendix 2. A total of 46 studies were further identified from these programs; most of them conducted a concurrent comparison (n=29) and used a passive comparator (n=39). The study characteristics are provided in detail in Multimedia Appendix 3.

Results of Primary Analyses

Unprotected Anal Intercourse

In combing all 35 studies that tested for UAI, a significant and small overall intervention effect was observed at end point (d_+=−.32; *P*<.001), albeit with considerable heterogeneity (I^2=83.3%). One study included subjects who reported recent condom-less sex with partners of either gender before enrollment; such a high-risk preintervention status might have contributed to the greatly decreased UAI (*d*=−1.13; SE 0.21) [49]. Moreover, the sample had the highest mean age of 45.1 years among all studies. Another study was focused on MSM sex workers and primarily addressed the context of transactional sex [63]. An extremely positive effect was found for reduced UAI with nonpaying male partners (*d*=−3.95; SE 0.35). Removing both outliers resulted in a smaller effect with moderate heterogeneity (d_+=−.21; *P*<.001; I^2=50.8%).

When stratified by the partner-specific outcome, a larger end point d_+ was shown for UAI with main partners than with nonmain partners. When stratified by the follow-up period, a significant d_+ was obtained at all points. The nested subgroup analysis found a significantly greater intervention effect when compared with preintervention groups than with concurrent groups (Q_b=13.38; *P*<.001); hence, the meta-analysis was performed separately for the 2 types of comparison status as recommended (Figures 2 and 3) [89].

Figure 2. A forest plot of intervention effects on unprotected anal intercourse against concurrent comparison groups.

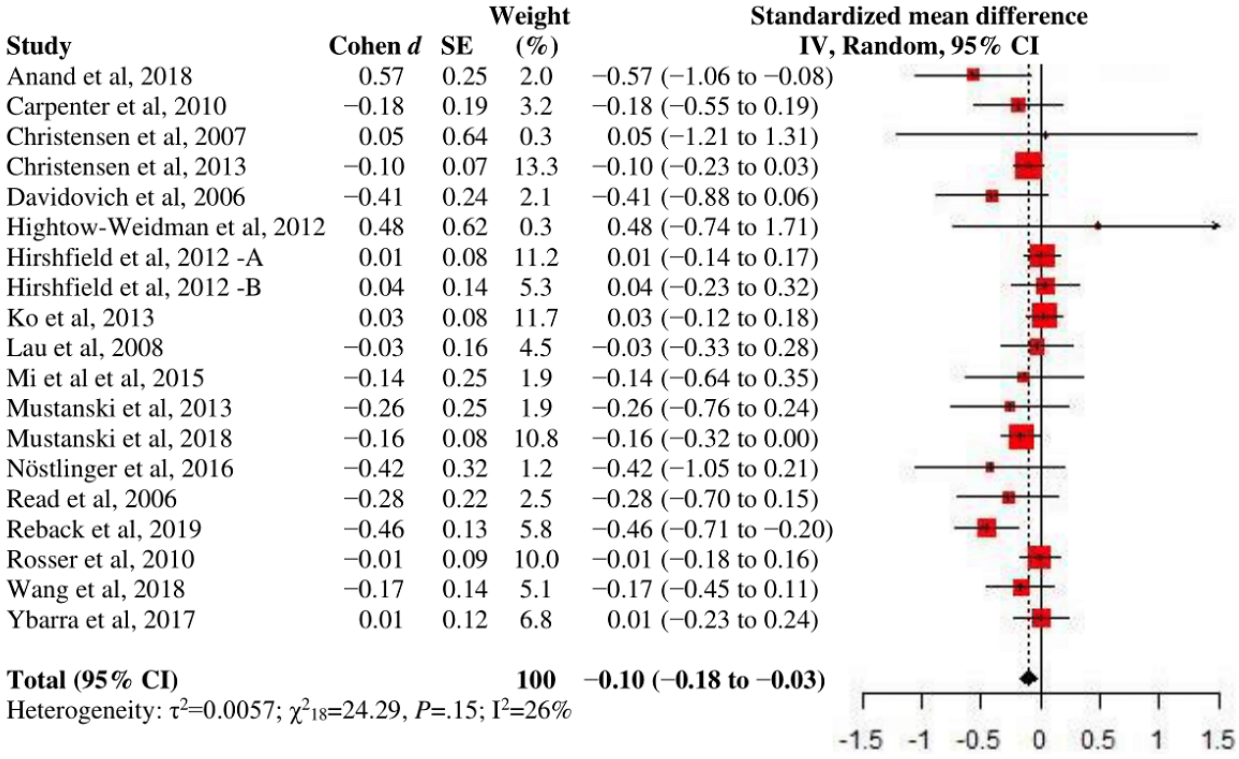
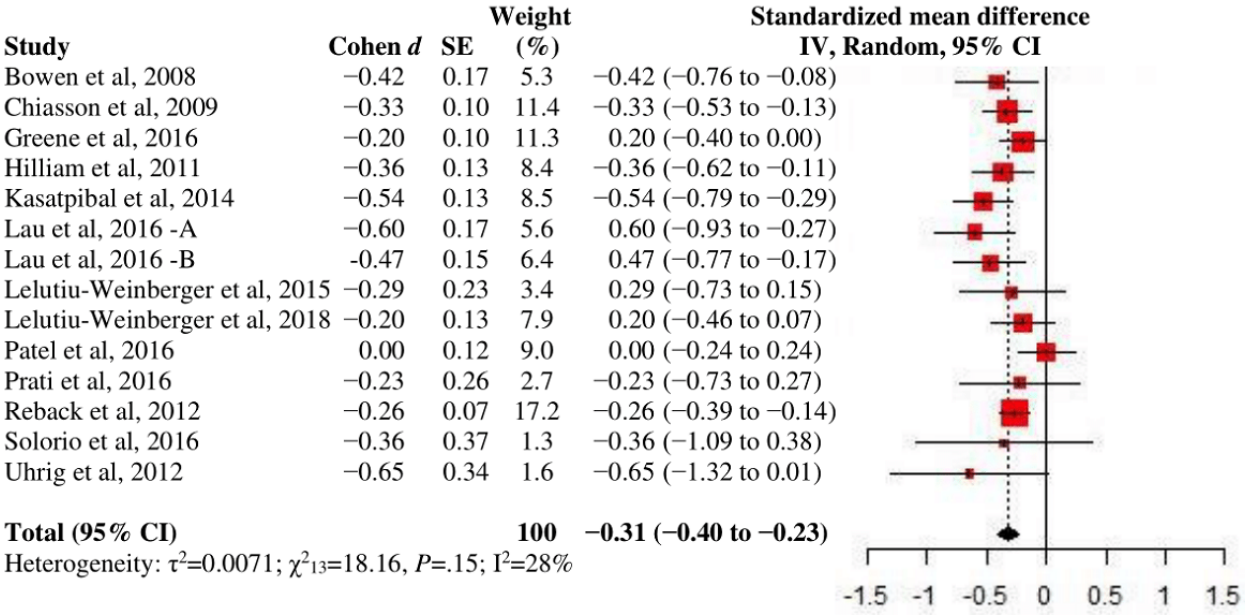


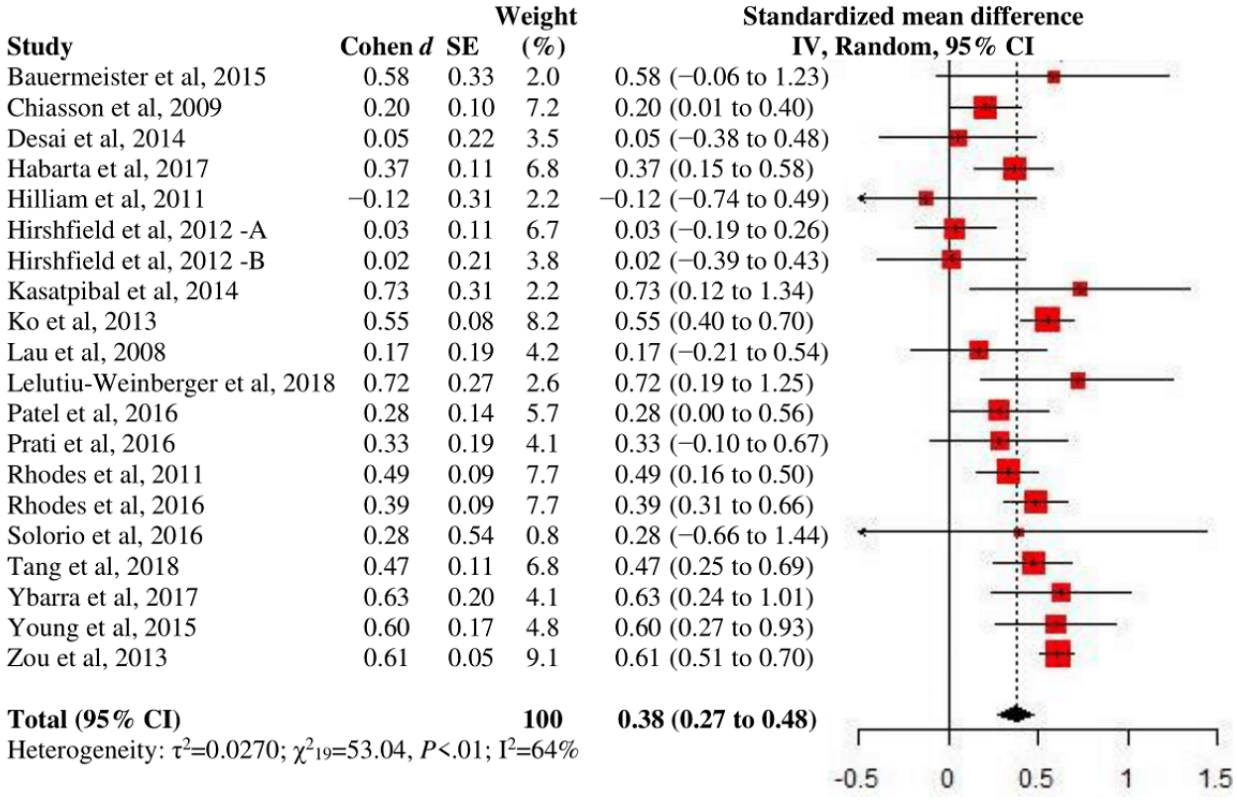
Figure 3. A forest plot of intervention effects on unprotected anal intercourse against preintervention comparison groups.



HIV Testing

Similarly, pooling the studies that tested for HIV testing revealed a significant, small overall effect at endpoint (n=23; d_+=−.32; P<.001). The considerable heterogeneity (I^2=84.0%) identified was largely attributable to the detection of outliers. Bourne et al [40] tested a texting reminder intervention against a blank control group involving participants who had refused such reminders; thus, the extreme effect size (d=.80; SE 0.06) could have been susceptible to volunteer bias. Wang et al [79] promoted a novel self-testing approach and distributed free home-based testing kits; the uptake of any type of testing was found to be considerably higher in the intervention group (d=1.18; SE 0.14). Mikolajczak et al [62] unexpectedly found a negative but nonsignificant intervention effect on HIV or STI testing relative to an active comparator (d=−.13; SE 0.12). The exclusion of outliers caused a slight change in the effectiveness but substantially reduced heterogeneity (d_+=.38; P<.001; I^2=64.2%; Figure 4). A larger d_+ was further shown at postintervention than at short-term follow-up; however, only one sample was followed up for more than 3 months. No significant group difference by comparison status was observed.

Figure 4. A forest plot of intervention effects on HIV testing.



Multiple Sex Partnership

A significant, small overall effect was observed on the reduced MSP at endpoint with moderate heterogeneity (d_+=−.26; P=.02; n=6; I^2=60.2%; Figure 5). The largest d_+ was shown at medium-term follow-up among all follow-up periods, although only a few studies were available for each grouping. No between-group test was performed for comparison status because of the small number of studies. Publication bias was detected for none of the outcomes. More results of primary analyses are presented in Table 2.

Figure 5. A forest plot of intervention effects on multiple sex partnership.

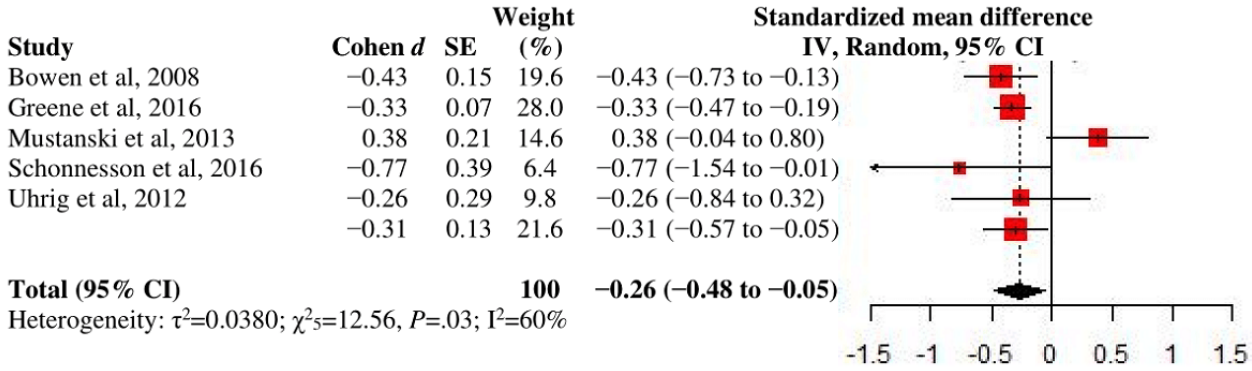


Table 2. Overall effect sizes and stratification by study design features.

Studies	Population, n	Cohen *d* (95% CI)	*P* value	Q_w (*P* value)[a]	I^2 (%)	Q_b (*P* value[b])
UAI[c]						N/A[d]
All studies at endpoint	35	−.32 (−.43 to −.20)	<.001	203.32 (<.001)	83.3	
Studies at endpoint without outliers	33	−.21 (−.28 to −.14)	<.001	65.08 (<.001)	50.8	
Partner-specific UAI[e,f,g]						N/A
With main partners	6	−.42 (−.57 to −.27)	<.001	5.18 (.39)	3.4	
With nonmain partners	9	−.20 (−.27 to −.13)	<.001	7.78 (.46)	0.0	
Follow-up period[f]						N/A
Postintervention	16	−.22 (−.36 to −.08)	.001	48.02 (<.001)	68.8	
≤3 months[e]	19	−.21 (−.28 to −.12)	<.001	35.92 (.007)	49.9	
>3 months[e]	6	−.26 (−.43 to −.08)	.004	11.60 (.04)	56.9	
Comparison status[e,g]						13.38 (<.001)
Concurrent	19	−.10 (−.18 to −.03)	.004	24.29 (.14)	25.9	
Preintervention	14	−.31 (−.40 to −.23)	<.001	18.16 (.15)	28.4	
HIV testing						
All studies at end point	23	.41 (.28 to .54)	<.001	137.43 (<.001)	84.0	
Studies at end point without outliers	20	.38 (.27 to .48)	<.001	53.04 (<.001)	64.2	
Follow-up period[f]						N/A
Postintervention	15	.46 (.38 to .55)	<.001	23.49 (.05)	40.4	
≤3 months[e]	8	.24 (.09 to .40)	.002	12.57 (.08)	44.3	
>3 months	1	.47 (.25 to .69)	<.001	N/A	N/A	
Comparison status[e,g]						0.42 (.52)
Concurrent	11	.39 (.25 to .53)	<.001	36.67 (<.001)	72.7	
Preintervention	9	.33 (.22 to .44)	<.001	9.14 (.33)	12.4	
Multiple sex partnership						
All studies at endpoint	6	−.26 (−.48 to −.05)	.02	12.56 (.03)	60.2	
Follow-up period[f]						N/A
Postintervention	2	−.18 (−.33 to −.03)	.02	0.08 (.78)	0.0	
≤3 months	3	−.19 (−.75 to .37)	.50	11.65 (.003)	82.8	
>3 months	2	−.36 (−.56 to −.16)	<.001	0.34 (.56)	0.0	

[a]Q_w denotes the degree of within-group heterogeneity.

[b]Q_b denotes the degree of between-group difference in the pooled effect sizes.

[c]UAI: unprotected anal intercourse.

[d]N/A: not applicable.

[e]Studies with exclusion of outliers.

[f]Non-nested groupings: one study could contribute to more than one grouping.

[g]Effect sizes at the endpoint were pooled.

Results of Moderation Analyses

Behavioral Treatment Components

Most of the studies were designed based on a theory (n=37); the most common one was the information-motivation-behavioral skills model (IMB; n=14). The intervention effects on UAI and HIV testing appeared to be comparable regardless of whether any theory was used. Among the theory-based studies, IMB was significantly associated with greater UAI reductions than other theories for concurrent comparisons (Q_b=4.33; *P*=.04). The more extensive

use of tailoring strategies was significantly associated with decreased UAI for concurrent comparisons (β=−.08; *P*=.01). Specifically, feedback was the only strategy that had a significantly negative effect on UAI (Q_b=4.00; *P*=.04), whereas adopting the personalization strategy tended to have a positive effect on HIV testing (Q_b=2.76; *P*=.10).

Significantly greater UAI reductions were shown for the tunneled (Q_b=7.23; *P*=.01) and self-paced (Q_b=4.23; *P*=.04) treatments relative to concurrent and preintervention groups, respectively. Such a moderating effect of navigation style was not detected for HIV testing. Longer treatments showed a significantly greater increase in HIV testing than those with a single session or those lasting ≤1 month (Q_b=16.97; *P*<.001). However, the effects on UAI did not significantly differ across treatment durations.

Electronic Health Technology Components

Over half of the interventions were operated via HCI (n=26), and the others were in a CMC (n=10) or combined (n=8) mode. More studies used a single modality (n=32) than those incorporating multi-modalities in the same intervention albeit with varying combination patterns (n=14). No between-group difference was observed in the intervention effects on UAI by operation mode or modality type. The CMC and combined modes tended to present a larger d_+ value for HIV testing than did the HCI mode (Q_b=2.76; *P*=.09). When separately exploring the effect of specific modes, the use of CMC predicted a significantly greater increase in HIV testing (Q_b=4.38; *P*=.04), whereas HCI use showed no association with effectiveness. Moreover, a significantly larger d_+ value was found for the use of messaging, social media, and multi-modalities than for static sites (Q_b=12.79; *P*=.005).

Intervention Adherence

Among all interventions, the adherence rates ranged from 25.5% to 100%. A higher adherence rate was significantly associated with a decrease in UAI for concurrent comparisons (β=−.27; *P*=.03). Given the highly negative skewed distribution, the studies were evenly divided into 3 levels to reflect the degree of relative adherence; the adherence rates were above 91.6% in the first tertile and below 76% in the third tertile. Intervention effects on HIV testing significantly differed by adherence level (Q_b=7.28; *P*=.03). A moderate d_+ value was observed for high-level studies (in the first tertile) and a small one for mid- and low-level studies. Details of the intervention features are presented in Multimedia Appendix 4; more results of moderation analyses are presented in Multimedia Appendix 5.

Discussion

Principal Findings

This study identified 46 eligible studies published since 2006, which highlight the increasing application of eHealth technology in intervention delivery for HIV prevention over the last decade. Most studies tested the effectiveness of an entire intervention against a passive comparator. Those with an active comparator delivered a treatment that was either a basic component of a comprehensive intervention package [79] or existing online information that the intervention group might be exposed to as well [62]; hence, it is possible to isolate the effects of specific intervention components.

Effectiveness of Electronic Health–Based Interventions

The primary meta-analysis consistently revealed a significant and small overall intervention effect for all the behavioral outcomes at endpoint. Considerable heterogeneity was nevertheless detected for UAI and HIV testing, and some influential cases were further identified to show extreme effect sizes and distinctive study characteristics. Removal of the outliers weakened the pooled effectiveness for both outcomes to some extent. In line with extant reviews, eHealth-based interventions exerted a greater impact on HIV testing [12] and the number of sex partners [14] than on condom use. Notably, the magnitude of effectiveness observed is seemingly higher than that in the general population [12], which suggests that the MSM community may benefit more than others from eHealth-based interventions.

Furthermore, the meta-analysis within the groupings by outcome type demonstrated greater reductions in UAI with main partners than with nonmain partners. It is recommended that future interventions target the partner-specific determinants of risky sexual behaviors. Most of the studies evaluated the postintervention and short-term effects across outcomes. Only 6 study samples were followed up for more than 3 months and showed a comparable decrease in UAI to that at earlier timepoints. It was not possible to estimate the long-term effect on HIV testing and MSP owing to the lack of data. This finding emphasizes the need to explore the effectiveness of eHealth interventions in maintaining behavior changes, especially given the necessity for consistent condom use and regular HIV testing [7,98]. The greater effect on UAI found for the pre-post comparison was probably confounded by factors unrelated to the intervention (eg, fatigue). Nevertheless, the heterogeneity between comparison types may also have resulted from the different intervention components deployed by the 2 groups of studies, as has been discussed in the following sections.

Predictors of Intervention Effectiveness

Behavioral Treatment Components

This study confirmed the previous finding [14] that the use of the theory did not moderate the effectiveness of eHealth-based interventions for HIV prevention, although the majority of treatments were developed on a theoretical basis. The overall extent of theory use failed to significantly influence effectiveness across outcomes. However, this finding may merely reflect a lack of sensitivity in distinguishing an effect [99]. The number of tailoring strategies used significantly moderated effectiveness in reducing UAI, and only the most frequently used strategy, that is, feedback, was further shown to be effective. Treatments were less tailored for interventions promoting HIV testing, and only the moderating effect of personalization use reached marginal significance. Some evidence supporting the superiority of certain strategies has been reported [23]. The use of feedback and personalization commonly featured tailoring at an individual level, whereas the

use of adaptation generated content matched to group-level factors.

Inconclusive results were found for the role of navigation style. Tunneled treatments achieved greater UAI reductions than self-paced treatments in studies using a concurrent group but lower reductions in those using a preintervention group. A plausible explanation is that the tunneling pattern differed across comparison types. A tunneled treatment by definition requires multiple interactions with the intervention, which could impair the effectiveness of eHealth behavioral interventions [23]. Over 60% of treatments with a preintervention group involved more than 5 modules, whereas only 24% of those with a concurrent group did so. Above all, the efficacy of the tunneling design itself is controversial as, on the one hand, the sequential release could ameliorate information anxiety and enhance the behavioral change process, whereas, on the other hand, the artificial confines could inhibit typical information-foraging behavior and intervention participants may lack the motivation to accommodate such constraints [100,101].

Longer treatments significantly predicted higher uptake of HIV testing but not reduced UAI, which adds to the evidence in favor of a difference in the influence of treatment duration on sexual and detection behaviors [12]. However, this finding might merely reject a linear relationship between treatment duration and the effect on UAI. A recent meta-analysis even demonstrated a negative effect of increased intervention length on intervention adherence and behavioral impacts, driven by decreased motivation over time [23].

Electronic Health Technology Components

Taking advantage of both human communication (eg, flexibility and rapport building [49]) and eHealth-enabled capacities (eg, convenience and anonymity [30]), the use of CMC was shown to be effective in increasing HIV testing. There was little evidence for the moderating effect of the operation mode on UAI, although only 1 intervention used a CMC mode. The results illustrate the need to incorporate human involvement into eHealth interventions. A variety of communicative functions have proved their efficacy in promoting behavior change, including counselor- or user-initiated conversations and peer-to-peer interactions [22]. Moreover, the greatest effectiveness was obtained with a combined mode, so the relevant question is not whether HCI or CMC is superior but, rather, how to combine them to maximize persuasiveness.

Similarly, intervention effects on the outcome of HIV testing but not UAI varied by modality type. Messaging, social media, and the combined type achieved a comparably small-to-medium effect. Messaging, largely operated via portable devices, affords individualized interventions (eg, location-based services) and dynamic connections (eg, response on request) [11,102]. Social media stands out for its capabilities to influence social norms and create a peer-supportive environment [103], especially within a closely connected community such as MSM [35]. It also enables interventions to blend into users' daily lives by utilizing well-established platforms (eg, Facebook) [32]. The use of hybrid modalities may further generate a synergistic effect.

Intervention Adherence

Consistent with previous literature [23], a high degree of adherence to eHealth interventions predicted a protective effect on UAI and HIV testing. Assuming intention-to-treat as the motive to engage in an intervention, adherence is rooted in the properties of intervention design. Efficacious strategies (eg, self-monitoring), observable benefits (eg, health outcomes), and feasible programming (eg, appropriate workload) can all bolster engagement [26,36]. Usability of eHealth technology is also critical to minimizing discontinuance [36].

Limitations

This study focused on the behavioral aspects of HIV prevention. Notably, it did not indicate the superiority of stand-alone behavioral interventions over other prevention tools (eg, pre-exposure prophylaxis) but was, rather, an attempt to explore the possibility of leveraging technological advances to strengthen an integral part of comprehensive biobehavioral prevention efforts, especially at a transition stage when the role of biomedical prevention is limited by its slow and uneven scale-up [7,104]. It is also impossible to synthesize evidence for other preventive behaviors owing to insufficient data (eg, serosorting) or highly heterogeneous measurements (eg, substance use). The potential interplay among intervention components has not been examined because of the lack of statistical power. Some other effectiveness predictors may also be missed (eg, eHealth literacy [26]), although the key factors with recognized terminology and accessible coding sources were selected. Finally, these findings could not represent studies reported in a language other than English, although there was no evidence of biased effectiveness among previous language-restrictive meta-analyses [105].

Conclusions

This meta-analysis demonstrated eHealth technology to be a promising tool for delivering HIV prevention interventions among the MSM population. Nevertheless, there is limited evidence on the long-term impact of such interventions, and more research is warranted to investigate their application in non-Western contexts. Our findings suggested that enhanced behavioral treatment (eg, the use of tailoring strategies) determined the effect of eHealth interventions on UAI reductions, whereas appropriate programming (eg, longer treatment duration) and advanced eHealth technology (eg, the use of social media) predicted an increase in HIV testing. Future intervention design should focus on integrating efficacious treatment and technology components as well as on exploring their potential interplay. It is also recommended to incorporate implementation strategies to improve intervention adherence and, thus, achieve better effectiveness.

Acknowledgments

The authors would like to express their heartfelt gratitude to Ms Beizhu Ye and Ms Mei Zhao, who provided support with screening, and to the authors of the original studies, who provided additional information necessary for this review. This study was supported by the Early Career Scheme for 2017-2018 (reference no. 24106317), and the funding agent was the University Grants Committee of Hong Kong. The funding source(s) was not involved in the literature review, systematic review, meta-analysis, and writing of the report.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The search strategy.
[DOCX File , 43 KB - jmir_v22i5e15977_app1.docx]

Multimedia Appendix 2
Results of quality assessment.
[DOCX File , 33 KB - jmir_v22i5e15977_app2.docx]

Multimedia Appendix 3
Characteristics of 46 studies included in the present meta-analysis.
[DOCX File , 153 KB - jmir_v22i5e15977_app3.docx]

Multimedia Appendix 4
Design and implementation features of included interventions.
[DOCX File , 156 KB - jmir_v22i5e15977_app4.docx]

Multimedia Appendix 5
Results of moderation analyses.
[DOCX File , 36 KB - jmir_v22i5e15977_app5.docx]

References

1. Beyrer C, Baral SD, van Griensven F, Goodreau SM, Chariyalertsak S, Wirtz AL, et al. Global epidemiology of HIV infection in men who have sex with men. Lancet 2012 Jul 28;380(9839):367-377 [FREE Full text] [doi: 10.1016/S0140-6736(12)60821-6] [Medline: 22819660]
2. Beyrer C, Sullivan P, Sanchez J, Baral SD, Collins C, Wirtz AL, et al. The increase in global HIV epidemics in MSM. AIDS 2013 Nov 13;27(17):2665-2678. [doi: 10.1097/01.aids.0000432449.30239.fe] [Medline: 23842129]
3. Hess KL, Crepaz N, Rose C, Purcell D, Paz-Bailey G. Trends in sexual behavior among men who have sex with men (MSM) in high-income countries, 1990-2013: a systematic review. AIDS Behav 2017 Oct;21(10):2811-2834 [FREE Full text] [doi: 10.1007/s10461-017-1799-1] [Medline: 28555317]
4. van Griensven F, de Lind van Wijngaarden JW. A review of the epidemiology of HIV infection and prevention responses among MSM in Asia. AIDS 2010 Sep;24(Suppl 3):S30-S40. [doi: 10.1097/01.aids.0000390087.22565.b4] [Medline: 20926926]
5. Mei S, Quax R, van de Vijver D, Zhu Y, Sloot PM. Increasing risk behaviour can outweigh the benefits of antiretroviral drug treatment on the HIV incidence among men-having-sex-with-men in Amsterdam. BMC Infect Dis 2011 May 11;11:118 [FREE Full text] [doi: 10.1186/1471-2334-11-118] [Medline: 21569307]
6. Campbell CK, Lippman SA, Moss N, Lightfoot M. Strategies to increase HIV testing among MSM: a synthesis of the literature. AIDS Behav 2018 Aug;22(8):2387-2412. [doi: 10.1007/s10461-018-2083-8] [Medline: 29550941]
7. Beyrer C, Baral SD, Collins C, Richardson ET, Sullivan PS, Sanchez J, et al. The global response to HIV in men who have sex with men. Lancet 2016 Jul 9;388(10040):198-206. [doi: 10.1016/S0140-6736(16)30781-4] [Medline: 27411880]
8. Higa DH, Crepaz N, Marshall KJ, Kay L, Vosburgh HW, Spikes P, et al. A systematic review to identify challenges of demonstrating efficacy of HIV behavioral interventions for gay, bisexual, and other men who have sex with men (MSM). AIDS Behav 2013 May;17(4):1231-1244 [FREE Full text] [doi: 10.1007/s10461-013-0418-z] [Medline: 23397183]
9. Herbst JH, Sherba RT, Crepaz N, Deluca JB, Zohrabyan L, Stall RD, HIV/AIDS Prevention Research Synthesis Team. A meta-analytic review of HIV behavioral interventions for reducing sexual risk behavior of men who have sex with men. J Acquir Immune Defic Syndr 2005 Jun 1;39(2):228-241. [Medline: 15905741]

10. Lu H, Liu Y, Dahiya K, Qian H, Fan W, Zhang L, et al. Effectiveness of HIV risk reduction interventions among men who have sex with men in China: a systematic review and meta-analysis. PLoS One 2013;8(8):e72747 [FREE Full text] [doi: 10.1371/journal.pone.0072747] [Medline: 24137497]
11. Noar SM, Willoughby JF. eHealth interventions for HIV prevention. AIDS Care 2012;24(8):945-952 [FREE Full text] [doi: 10.1080/09540121.2012.668167] [Medline: 22519523]
12. Swanton R, Allom V, Mullan B. A meta-analysis of the effect of new-media interventions on sexual-health behaviours. Sex Transm Infect 2015 Feb;91(1):14-20. [doi: 10.1136/sextrans-2014-051743] [Medline: 25433051]
13. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
14. Noar SM, Black HG, Pierce LB. Efficacy of computer technology-based HIV prevention interventions: a meta-analysis. AIDS 2009 Jan 2;23(1):107-115. [doi: 10.1097/QAD.0b013e32831c5500] [Medline: 19050392]
15. Noar SM, Pierce LB, Black HG. Can computer-mediated interventions change theoretical mediators of safer sex? A meta-analysis. Hum Commun Res 2010;36(3):261-297. [doi: 10.1111/j.1468-2958.2010.01376.x]
16. Holloway IW, Rice E, Gibbs J, Winetrobe H, Dunlap S, Rhoades H. Acceptability of smartphone application-based HIV prevention among young men who have sex with men. AIDS Behav 2014 Feb;18(2):285-296 [FREE Full text] [doi: 10.1007/s10461-013-0671-1] [Medline: 24292281]
17. Zou HC, Fan S. Characteristics of men who have sex with men who use smartphone geosocial networking applications and implications for HIV interventions: a systematic review and meta-analysis. Arch Sex Behav 2017 May;46(4):885-894. [doi: 10.1007/s10508-016-0709-3] [Medline: 27043837]
18. Knight R, Karamouzian M, Salway T, Gilbert M, Shoveller J. Online interventions to address HIV and other sexually transmitted and blood-borne infections among young gay, bisexual and other men who have sex with men: a systematic review. J Int AIDS Soc 2017 Nov;20(3):- [FREE Full text] [doi: 10.1002/jia2.25017] [Medline: 29091340]
19. Schnall R, Travers J, Rojas M, Carballo-Diéguez A. eHealth interventions for HIV prevention in high-risk men who have sex with men: a systematic review. J Med Internet Res 2014 May 26;16(5):e134 [FREE Full text] [doi: 10.2196/jmir.3393] [Medline: 24862459]
20. Deeks JJ, Higgins JP, Altman DG. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0.: The Cochrane Collaboration; 2011. 9.6.5.2 Specify Characteristics in Advance URL: https://handbook-5-1.cochrane.org/chapter_9/9_6_5_2_specify_characteristics_in_advance.htm [accessed 2018-09-01]
21. Michie S, Prestwich A. Are interventions theory-based? Development of a theory coding scheme. Health Psychol 2010 Jan;29(1):1-8. [doi: 10.1037/a0016939] [Medline: 20063930]
22. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
23. Cugelman B, Thelwall M, Dawes P. Online interventions for social marketing health behavior change campaigns: a meta-analysis of psychological architectures and adherence factors. J Med Internet Res 2011 Feb 14;13(1):e17 [FREE Full text] [doi: 10.2196/jmir.1367] [Medline: 21320854]
24. Lustria ML, Cortese J, Noar SM, Glueckauf RL. Computer-tailored health interventions delivered over the Web: review and analysis of key components. Patient Educ Couns 2009 Feb;74(2):156-173. [doi: 10.1016/j.pec.2008.08.023] [Medline: 18947966]
25. Kelders SM, Kok RN, Ossebaard HC, van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
26. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
27. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun 5;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
28. Wahle F, Bollhalder L, Kowatsch T, Fleisch E. Toward the design of evidence-based mental health information systems for people with depression: a systematic literature review and meta-analysis. J Med Internet Res 2017 May 31;19(5):e191 [FREE Full text] [doi: 10.2196/jmir.7381] [Medline: 28566267]
29. Wildeboer G, Kelders SM, van Gemert-Pijnen JE. The relationship between persuasive technology principles, adherence and effect of web-Based interventions for mental health: A meta-analysis. Int J Med Inform 2016 Dec;96:71-85. [doi: 10.1016/j.ijmedinf.2016.04.005] [Medline: 27117057]
30. Barak A, Hen L, Boniel-Nissim M, Shapira N. A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. J Technol Hum Serv 2008;26(2-4):109-160. [doi: 10.1080/15228830802094429]

31. Kuester A, Niemeyer H, Knaevelsrud C. Internet-based interventions for posttraumatic stress: a meta-analysis of randomized controlled trials. Clin Psychol Rev 2016 Feb;43:1-16. [doi: 10.1016/j.cpr.2015.11.004] [Medline: 26655959]
32. Muessig KE, Nekkanti M, Bauermeister J, Bull S, Hightow-Weidman LB. A systematic review of recent smartphone, internet and web 2.0 interventions to address the HIV continuum of care. Curr HIV/AIDS Rep 2015 Mar;12(1):173-190 [FREE Full text] [doi: 10.1007/s11904-014-0239-3] [Medline: 25626718]
33. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015 Mar 18;36:393-415 [FREE Full text] [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
34. Head KJ, Noar SM, Iannarino NT, Harrington NG. Efficacy of text messaging-based interventions for health promotion: a meta-analysis. Soc Sci Med 2013 Nov;97:41-48. [doi: 10.1016/j.socscimed.2013.08.003] [Medline: 24161087]
35. Cao B, Gupta S, Wang J, Hightow-Weidman LB, Muessig KE, Tang W, et al. Social media interventions to promote HIV testing, linkage, adherence, and retention: systematic review and meta-analysis. J Med Internet Res 2017 Nov 24;19(11):e394 [FREE Full text] [doi: 10.2196/jmir.7997] [Medline: 29175811]
36. 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]
37. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009 Jul 21;6(7):e1000100 [FREE Full text] [doi: 10.1371/journal.pmed.1000100] [Medline: 19621070]
38. Anand T, Nitpolprasert C, Jantarapakde J, Meksena R, Phomthong S, Phoseeta P, et al. Implementation and impact of a technology-based HIV risk-reduction intervention among Thai men who have sex with men using 'Vialogues': a randomized controlled trial. AIDS Care 2020 Mar;32(3):394-405. [doi: 10.1080/09540121.2019.1622638] [Medline: 31154811]
39. Bauermeister JA, Pingel ES, Jadwin-Cakmak L, Harper GW, Horvath K, Weiss G, et al. Acceptability and preliminary efficacy of a tailored online HIV/STI testing intervention for young men who have sex with men: the Get Connected! program. AIDS Behav 2015 Oct;19(10):1860-1874 [FREE Full text] [doi: 10.1007/s10461-015-1009-y] [Medline: 25638038]
40. Bourne C, Knight V, Guy R, Wand H, Lu H, McNulty A. Short message service reminder intervention doubles sexually transmitted infection/HIV re-testing rates among men who have sex with men. Sex Transm Infect 2011 Apr;87(3):229-231. [doi: 10.1136/sti.2010.048397] [Medline: 21296796]
41. Bowen AM, Williams ML, Daniel CM, Clayton S. Internet based HIV prevention research targeting rural MSM: feasibility, acceptability, and preliminary efficacy. J Behav Med 2008 Dec;31(6):463-477 [FREE Full text] [doi: 10.1007/s10865-008-9171-6] [Medline: 18770021]
42. Carpenter KM, Stoner SA, Mikko AN, Dhanak LP, Parsons JT. Efficacy of a web-based intervention to reduce sexual risk in men who have sex with men. AIDS Behav 2010 Jun;14(3):549-557 [FREE Full text] [doi: 10.1007/s10461-009-9578-2] [Medline: 19499321]
43. Chiasson MA, Shaw FS, Humberstone M, Hirshfield S, Hartel D. Increased HIV disclosure three months after an online video intervention for men who have sex with men (MSM). AIDS Care 2009 Sep;21(9):1081-1089. [doi: 10.1080/09540120902730013] [Medline: 20024766]
44. Christensen JL. CORE – Aggregating the world's open access research papers. Ann Arbor: University of Southern California; 2007. When It's Good to Feel Bad: How Responses to Virtual Environments Predict Real-Life Sexual Risk-reduction URL: https://core.ac.uk/display/69137530 [accessed 2018-09-01]
45. Christensen JL, Miller LC, Appleby PR, Corsbie-Massay C, Godoy CG, Marsella SC, et al. Reducing shame in a game that predicts HIV risk reduction for young adult MSM: a randomized trial delivered nationally over the Web. J Int AIDS Soc 2013 Nov 13;16(3 Suppl 2):18716 [FREE Full text] [doi: 10.7448/IAS.16.3.18716] [Medline: 24242264]
46. Daniel C, Bowen A, Williams M, Clayton S, Ross L. Assessment of long-term efficacy of an internet delivered HIV risk reduction intervention for rural MSM. In: Abstracts of the 29th Annual Meeting and Scientific Sessions of the Society of Behavioral Medicine.: Annals of behavioral medicine; 2008 Presented at: Behavioral medicine: celebrating the past, inspiring the future; March 26-29, 2008; San Diego p. s105 URL: https://academic.oup.com/abm/article/35/suppl_1/s1/4569634
47. Davidovich U, de Wit JBF, Stroebe W. Using the Internet to reduce risk of HIV-infection in steady relationships: a randomized controlled trial of a tailored intervention for gay men. In: Davidovich U, editor. Liaisons Dangereuses: HIV Risk Behavior and Prevention in Steady Gay Relationships. Amsterdam: Roel & Uitgeefprojecten; 2006:96-122.
48. Desai M, Nardone A, Burns F, Mercey D, Gilson R, Muniina P, et al. Active recall of men who have sex with men (MSM) for an HIV/STI testing: a feasible and effective strategy? In: Abstracts of the Third Joint Conference of the British HIV Association (BHIVA) with the British Association for Sexual Health and HIV (BASHH). 2014 Presented at: Third Joint Conference of the BHIVA and the BASHH; April 1-4, 2014; Liverpool p. 109 URL: https://onlinelibrary.wiley.com/doi/10.1111/hiv.12147
49. Fernandez MI, Hosek SG, Hotton AL, Gaylord SE, Hernandez N, Alfonso SV, et al. A Randomized Controlled Trial of POWER: an internet-based HIV prevention intervention for black bisexual men. AIDS Behav 2016 Sep;20(9):1951-1960. [doi: 10.1007/s10461-016-1403-0] [Medline: 27085548]
50. Greene GJ, Madkins K, Andrews K, Dispenza J, Mustanski B. Implementation and Evaluation of the Keep It Up! Online HIV Prevention Intervention in a Community-Based Setting. AIDS Educ Prev 2016 Jun;28(3):231-245. [doi: 10.1521/aeap.2016.28.3.231] [Medline: 27244191]

51. Habarta N, Boudewyns V, Badal H, Johnston J, Uhrig J, Green D, et al. CDC'S Testing Makes Us Stronger (TMUS) Campaign: Was Campaign Exposure Associated With HIV Testing Behavior Among Black Gay and Bisexual Men? AIDS Educ Prev 2017 Jun;29(3):228-240. [doi: 10.1521/aeap.2017.29.3.228] [Medline: 28650228]
52. Hightow-Weidman LB, Pike E, Fowler B, Matthews DM, Kibe J, McCoy R, et al. HealthMpowerment.org: feasibility and acceptability of delivering an internet intervention to young Black men who have sex with men. AIDS Care 2012;24(7):910-920 [FREE Full text] [doi: 10.1080/09540121.2011.647677] [Medline: 22272759]
53. Hilliam A, Fraser L, Turner L. NHS Health Scotland. Scotland: NHS Health Scotland; 2011 Feb. HIV Wake-Up Campaign Evaluation URL: http://www.healthscotland.com/uploads/documents/15963-HIV%20Wake%20Up%20Campaign%20Evaluation.pdf [accessed 2018-09-01]
54. Hirshfield S, Chiasson MA, Joseph H, Scheinmann R, Johnson WD, Remien RH, et al. An online randomized controlled trial evaluating HIV prevention digital media interventions for men who have sex with men. PLoS One 2012;7(10):e46252 [FREE Full text] [doi: 10.1371/journal.pone.0046252] [Medline: 23071551]
55. Kasatpibal N, Viseskul N, Srikantha W, Fongkaew W, Surapagdee N, Grimes RM. Effects of internet-based instruction on HIV-prevention knowledge and practices among men who have sex with men. Nurs Health Sci 2014 Dec;16(4):514-520 [FREE Full text] [doi: 10.1111/nhs.12135] [Medline: 24645824]
56. Ko N, Hsieh C, Wang M, Lee C, Chen C, Chung A, et al. Effects of Internet popular opinion leaders (iPOL) among internet-using men who have sex with men. J Med Internet Res 2013 Feb 25;15(2):e40 [FREE Full text] [doi: 10.2196/jmir.2264] [Medline: 23439583]
57. Lau JT, Lau M, Cheung A, Tsui HY. A randomized controlled study to evaluate the efficacy of an internet-based intervention in reducing HIV risk behaviors among men who have sex with men in Hong Kong. AIDS Care 2008 Aug;20(7):820-828. [doi: 10.1080/09540120701694048] [Medline: 18608057]
58. Lau JTF, Lee AL, Tse WS, Mo PK, Fong F, Wang Z, et al. A randomized control trial for evaluating efficacies of two online cognitive interventions with and without fear-appeal imagery approaches in preventing unprotected anal sex among Chinese men who have sex with men. AIDS Behav 2016 Sep;20(9):1851-1862. [doi: 10.1007/s10461-015-1263-z] [Medline: 26802004]
59. Lelu iu-Weinberger C, Manu M, Ionescu F, Dogaru B, Kovacs T, Dorobăn escu C, et al. An mHealth intervention to improve young gay and bisexual men's sexual, behavioral, and mental health in a structurally stigmatizing national context. JMIR Mhealth Uhealth 2018 Nov 14;6(11):e183 [FREE Full text] [doi: 10.2196/mhealth.9283] [Medline: 30429117]
60. Lelutiu-Weinberger C, Pachankis JE, Gamarel KE, Surace A, Golub SA, Parsons JT. Feasibility, acceptability, and preliminary efficacy of a live-chat social media intervention to reduce HIV risk among young men who have sex with men. AIDS Behav 2015 Jul;19(7):1214-1227 [FREE Full text] [doi: 10.1007/s10461-014-0911-z] [Medline: 25256808]
61. Mi G, Wu Z, Wang X, Shi CX, Yu F, Li T, et al. Effects of a quasi-randomized web-based intervention on risk behaviors and treatment seeking among HIV-positive men who have sex with men in Chengdu, China. Curr HIV Res 2015;13(6):490-496 [FREE Full text] [doi: 10.2174/1570162x13666150624104522] [Medline: 26105555]
62. Mikolajczak J, Breukelen van G, Kok G, Hospers H. Evaluation of an online HIV-prevention intervention to promote HIV-testing among men who have sex with men: a randomised controlled trial. Neth J Psychol 2012;67(2):21-35 [FREE Full text]
63. Mimiaga MJ, Thomas B, Biello K, Johnson BE, Swaminathan S, Navakodi P, et al. A pilot randomized controlled trial of an integrated in-person and mobile phone delivered counseling and text messaging intervention to reduce HIV transmission risk among male sex workers in Chennai, India. AIDS Behav 2017 Nov;21(11):3172-3181 [FREE Full text] [doi: 10.1007/s10461-017-1884-5] [Medline: 28831618]
64. Mustanski B, Garofalo R, Monahan C, Gratzer B, Andrews R. Feasibility, acceptability, and preliminary efficacy of an online HIV prevention program for diverse young men who have sex with men: the keep it up! intervention. AIDS Behav 2013 Nov;17(9):2999-3012 [FREE Full text] [doi: 10.1007/s10461-013-0507-z] [Medline: 23673793]
65. Mustanski B, Parsons JT, Sullivan PS, Madkins K, Rosenberg E, Swann G. Biomedical and Behavioral Outcomes of Keep It Up!: An eHealth HIV Prevention Program RCT. Am J Prev Med 2018 Aug;55(2):151-158 [FREE Full text] [doi: 10.1016/j.amepre.2018.04.026] [Medline: 29937115]
66. Nöstlinger C, Platteau T, Bogner J, Buyze J, Dec-Pietrowska J, Dias S, Eurosupport Study Group. Implementation and operational research: computer-assisted intervention for safer sex in HIV-positive men having sex with men: findings of a European randomized multi-center trial. J Acquir Immune Defic Syndr 2016 Mar 1;71(3):e63-e72 [FREE Full text] [doi: 10.1097/QAI.0000000000000882] [Medline: 26866955]
67. Patel VV, Rawat S, Lelutiu-Weinberger C, Dange A, Kamath C, Poojary R, et al. CHALO! A social media based peer-delivered intervention increases HIV testing in men who have sex with men in Mumbai, India: a randomized trial. In: Oral abstracts of the 21st International AIDS Conference.: Journal of the International AIDS Society; 2016 Presented at: the 21st International AIDS Conference; 18–22 July 2016; Durban p. 103 URL: https://onlinelibrary.wiley.com/doi/full/10.7448/IAS.19.6.21264
68. Prati G, Mazzoni D, Cicognani E, Albanesi C, Zani B. Evaluating the persuasiveness of an HIV mass communication campaign using gain-framed messages and aimed at creating a superordinate identity. Health Commun 2016 Sep;31(9):1097-1104. [doi: 10.1080/10410236.2015.1040983] [Medline: 26861894]

69. Read SJ, Miller LC, Appleby PR, Nwosu ME, Reynaldo S, Lauren A, et al. Socially optimized learning in a virtual environment: Reducing risky sexual behavior among men who have sex with men. Human Comm Res 2006 Jan;32(1):1-34. [doi: 10.1111/j.1468-2958.2006.00001.x]
70. Reback CJ, Fletcher JB, Swendeman DA, Metzner M. Theory-based text-messaging to reduce methamphetamine use and HIV sexual risk behaviors among men who have sex with men: automated unidirectional delivery outperforms bidirectional peer interactive delivery. AIDS Behav 2019 Jan;23(1):37-47 [FREE Full text] [doi: 10.1007/s10461-018-2225-z] [Medline: 30006792]
71. Reback CJ, Grant DL, Fletcher JB, Branson CM, Shoptaw S, Bowers JR, et al. Text messaging reduces HIV risk behaviors among methamphetamine-using men who have sex with men. AIDS Behav 2012 Oct;16(7):1993-2002 [FREE Full text] [doi: 10.1007/s10461-012-0200-7] [Medline: 22610370]
72. Rhodes SD, McCoy TP, Tanner AE, Stowers J, Bachmann LH, Nguyen AL, et al. Using social media to increase HIV testing among gay and bisexual men, other men who have sex with men, and transgender persons: Outcomes from a randomized community trial. Clin Infect Dis 2016 Jun 1;62(11):1450-1453 [FREE Full text] [doi: 10.1093/cid/ciw127] [Medline: 26980878]
73. Rhodes SD, Vissman AT, Stowers J, Miller C, McCoy TP, Hergenrather KC, et al. A CBPR partnership increases HIV testing among men who have sex with men (MSM): outcome findings from a pilot test of the CyBER/testing internet intervention. Health Educ Behav 2011 Jun;38(3):311-320 [FREE Full text] [doi: 10.1177/1090198110379572] [Medline: 21393625]
74. Rosser BR, Oakes JM, Konstan J, Hooper S, Horvath KJ, Danilenko GP, et al. Reducing HIV risk behavior of men who have sex with men through persuasive computing: results of the Men's INTernet Study-II. AIDS 2010 Aug 24;24(13):2099-2107 [FREE Full text] [doi: 10.1097/QAD.0b013e32833c4ac7] [Medline: 20601853]
75. Schonnesson LN, Bowen AM, Williams ML. Project SMART: Preliminary results from a test of the efficacy of a Swedish internet-based HIV risk-reduction intervention for men who have sex with men. Arch Sex Behav 2016 Aug;45(6):1501-1511. [doi: 10.1007/s10508-015-0608-z] [Medline: 26545912]
76. Solorio R, Norton-Shelpuk P, Forehand M, Montaño D, Stern J, Aguirre J, et al. Tu Amigo Pepe: Evaluation of a multi-media marketing campaign that targets young Latino immigrant MSM with HIV testing messages. AIDS Behav 2016 Sep;20(9):1973-1988. [doi: 10.1007/s10461-015-1277-6] [Medline: 26850101]
77. Tang W, Wei C, Cao B, Wu D, Li KT, Lu H, et al. Crowdsourcing to expand HIV testing among men who have sex with men in China: a closed cohort stepped wedge cluster randomized controlled trial. PLoS Med 2018 Aug;15(8):e1002645 [FREE Full text] [doi: 10.1371/journal.pmed.1002645] [Medline: 30153265]
78. Uhrig JD, Lewis MA, Bann CM, Harris JL, Furberg RD, Coomes CM, et al. Addressing HIV knowledge, risk reduction, social support, and patient involvement using SMS: results of a proof-of-concept study. J Health Commun 2012;17(Suppl 1):128-145. [doi: 10.1080/10810730.2011.649156] [Medline: 22548606]
79. Wang Z, Lau JT, Ip M, Ho SP, Mo PK, Latkin C, et al. A randomized controlled trial evaluating efficacy of promoting a home-based HIV self-testing with online counseling on increasing HIV testing among men who have sex with men. AIDS Behav 2018 Jan;22(1):190-201. [doi: 10.1007/s10461-017-1887-2] [Medline: 28831616]
80. Ybarra ML, Prescott TL, Phillips GL, Bull SS, Parsons JT, Mustanski B. Pilot RCT results of an mHealth HIV prevention program for sexual minority male adolescents. Pediatrics 2017 Jul;140(1):- [FREE Full text] [doi: 10.1542/peds.2016-2999] [Medline: 28659456]
81. Young SD, Cumberland WG, Nianogo R, Menacho LA, Galea JT, Coates T. The HOPE social media intervention for global HIV prevention in Peru: a cluster randomised controlled trial. Lancet HIV 2015 Jan;2(1):e27-e32 [FREE Full text] [doi: 10.1016/S2352-3018(14)00006-X] [Medline: 26236767]
82. Zou H, Fairley CK, Guy R, Bilardi J, Bradshaw CS, Garland SM, et al. Automated, computer generated reminders and increased detection of gonorrhoea, chlamydia and syphilis in men who have sex with men. PLoS One 2013;8(4):e61972 [FREE Full text] [doi: 10.1371/journal.pone.0061972] [Medline: 23613989]
83. Flores SA, Crepaz N, HIV Prevention Research Synthesis Team. Quality of study methods in individual- and group-level HIV intervention research: critical reporting elements. AIDS Educ Prev 2004 Aug;16(4):341-352. [doi: 10.1521/aeap.16.4.341.40396] [Medline: 15342336]
84. Higgins JP, Deeks JJ, Altman DG. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0.: The Cochrane Collaboration; 2011. 16.5 Studies With More Than Two Intervention Groups URL: https://handbook-5-1.cochrane.org/chapter_16/16_5_studies_with_more_than_two_intervention_groups.htm [accessed 2018-09-01]
85. Portnoy DB, Scott-Sheldon LA, Johnson BT, Carey MP. Computer-delivered interventions for health promotion and behavioral risk reduction: a meta-analysis of 75 randomized controlled trials, 1988-2007. Prev Med 2008 Jul;47(1):3-16 [FREE Full text] [doi: 10.1016/j.ypmed.2008.02.014] [Medline: 18403003]
86. Cushing CC, Steele RG. A meta-analytic review of eHealth interventions for pediatric health promoting and maintaining behaviors. J Pediatr Psychol 2010 Oct;35(9):937-949. [doi: 10.1093/jpepsy/jsq023] [Medline: 20392790]
87. Higgins JP, Deek JJ. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0.: The Cochrane Collaboration; 2011. 7.7.3.1 Post-Intervention Versus Change From Baseline URL: https://handbook-5-1.cochrane.org/chapter_7/7_7_3_1_post_intervention_versus_change_from_baseline.htm [accessed 2018-09-01]

88. Deeks JJ, Higgins JP, Altman DG. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0.: The Cochrane Collaboration; 2011. 9.2.3.2 The Standardized Mean Difference URL: https://handbook-5-1.cochrane.org/chapter_9/9_2_3_2_the_standardized_mean_difference.htm [accessed 2018-09-01]
89. Morris SB, DeShon RP. Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. Psychol Methods 2002 Mar;7(1):105-125. [doi: 10.1037/1082-989x.7.1.105] [Medline: 11928886]
90. Wilson DB. The Campbell Collaboration. Fairfax: George Mason University; 2001. Practical Meta-Analysis Effect Size Calculator URL: http://www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-Home.php [accessed 2018-09-01]
91. Lenhard W, Lenhard A. Psychometrica. Dettelbach: Psychometrica; 2016. Computation of Effect Sizes URL: http://www.psychometrica.de/effect_size.html [accessed 2018-09-01]
92. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Abingdon, United Kingdom: Routledge; 2013.
93. Deeks JJ, Higgins JP, Altman DG. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0.: The Cochrane Collaboration; 2011. 9.5.2 Identifying and Measuring Heterogeneity URL: https://handbook-5-1.cochrane.org/chapter_9/9_5_2_identifying_and_measuring_heterogeneity.htm [accessed 2018-09-01]
94. Viechtbauer W, Cheung MW. Outlier and influence diagnostics for meta-analysis. Res Synth Methods 2010 Apr;1(2):112-125. [doi: 10.1002/jrsm.11] [Medline: 26061377]
95. Sterne JA, Egger M, Moher D. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0.: The Cochrane Collaboration; 2011. 10.4.3.1 Recommendations on Testing for Funnel Plot Asymmetry URL: https://handbook-5-1.cochrane.org/chapter_10/10_4_3_1_recommendations_on_testing_for_funnel_plot_asymmetry.htm [accessed 2018-09-01]
96. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0.: The Cochrane Collaboration; 2011. 9.6.5.1 Ensure That There Are Adequate Studies to Justify Subgroup Analyses and Meta-Regressions URL: https://handbook-5-1.cochrane.org/chapter_9/9_6_5_1_ensure_that_there_are_adequate_studies_to_justify.htm
97. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. J Stat Softw 2010;36(3):1-48. [doi: 10.18637/jss.v036.i03]
98. Smith DK, Herbst JH, Zhang X, Rose CE. Condom effectiveness for HIV prevention by consistency of use among men who have sex with men in the United States. J Acquir Immune Defic Syndr 2015 Mar 1;68(3):337-344. [doi: 10.1097/QAI.0000000000000461] [Medline: 25469526]
99. Garnett C, Crane D, Brown J, Kaner E, Beyer F, Muirhead C, et al. Reported theory use by digital interventions for hazardous and harmful alcohol consumption, and association with effectiveness: meta-regression. J Med Internet Res 2018 Feb 28;20(2):e69 [FREE Full text] [doi: 10.2196/jmir.8807] [Medline: 29490895]
100. Danaher BG, McKay HG, Seeley JR. The information architecture of behavior change websites. J Med Internet Res 2005 May 18;7(2):e12 [FREE Full text] [doi: 10.2196/jmir.7.2.e12] [Medline: 15914459]
101. Lehto T, Oinas-Kukkonen H. Persuasive features in web-based alcohol and smoking interventions: a systematic review of the literature. J Med Internet Res 2011 Jul 22;13(3):e46 [FREE Full text] [doi: 10.2196/jmir.1559] [Medline: 21795238]
102. Reback CJ, Ling D, Shoptaw S, Rohde J. Developing a text messaging risk reduction intervention for methamphetamine-using MSM: research note. Open AIDS J 2010 May 14;4:116-122 [FREE Full text] [doi: 10.2174/1874613601004030116] [Medline: 20657827]
103. Tso LS, Tang W, Li H, Yan HY, Tucker JD. Social media interventions to prevent HIV: a review of interventions and methodological considerations. Curr Opin Psychol 2016 Jun 1;9:6-10 [FREE Full text] [doi: 10.1016/j.copsyc.2015.09.019] [Medline: 26516632]
104. Spinner CD, Boesecke C, Zink A, Jessen H, Stellbrink H, Rockstroh JK, et al. HIV pre-exposure prophylaxis (PrEP): a review of current knowledge of oral systemic HIV PrEP in humans. Infection 2016 Apr;44(2):151-158. [doi: 10.1007/s15010-015-0850-2] [Medline: 26471511]
105. Morrison A, Polisena J, Husereau D, Moulton K, Clark M, Fiander M, et al. The effect of English-language restriction on systematic review-based meta-analyses: a systematic review of empirical studies. Int J Technol Assess Health Care 2012 Apr;28(2):138-144. [doi: 10.1017/S0266462312000086] [Medline: 22559755]

Abbreviations

CMC: computer-mediated communication
eHealth: electronic health
HCI: human-computer interaction
IMB: information-motivation-behavioral skills model
MSM: men who have sex with men
MSP: multiple sex partnership
STI: sexually transmitted infection
UAI: unprotected anal intercourse

Edited by G Eysenbach; submitted 26.08.19; peer-reviewed by F Yu, Z Wang; comments to author 28.09.19; revised version received 19.12.19; accepted 01.02.20; published 25.05.20.

Please cite as:
Xin M, Viswanath K, Li AYC, CAO W, HU Y, Lau JTF, Mo PKH
The Effectiveness of Electronic Health Interventions for Promoting HIV-Preventive Behaviors Among Men Who Have Sex With Men: Meta-Analysis Based on an Integrative Framework of Design and Implementation Features
J Med Internet Res 2020;22(5):e15977
URL: https://www.jmir.org/2020/5/e15977
doi:10.2196/15977
PMID:32449685

Review

Effectiveness of Internet-Based Electronic Technology Interventions on Breastfeeding Outcomes: Systematic Review

Alaa Ali Almohanna[1], MSc; Khin Than Win[1], PhD; Shahla Meedya[2], PhD

[1]School of Computing and Information Technology, University of Wollongong, Wollongong, Australia

[2]School of Nursing, University of Wollongong, Wollongong, Australia

Corresponding Author:
Alaa Ali Almohanna, MSc
School of Computing and Information Technology
University of Wollongong
Northfields Ave
Wollongong, 2522
Australia
Phone: 61 402592326
Email: aaa933@uowmail.edu.au

Abstract

Background: Supporting women to initiate and continue breastfeeding is a global challenge. A range of breastfeeding interventions employing electronic technologies (e-technologies) are being developed, which offer different delivery modes and features over the internet; however, the impact of internet-based e-technologies on breastfeeding outcomes remains unclear.

Objective: This study aimed to identify the characteristics of current internet-based breastfeeding interventions employing e-technologies and investigate the effects of internet-based e-technologies on breastfeeding outcomes.

Methods: A systematic search was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines in the following databases: Scopus, Web of Science, the Cochrane Database of Systematic Reviews, ScienceDirect, Google Scholar, the Association for Computing Machinery, SpringerLink, and Institute of Electrical and Electronics Engineers Xplore.

Results: This systematic review included 16 studies published between 2007 and 2018, with 4018 women in 8 countries. The characteristics of the interventions were grouped based on (1) mode of delivery (web-based, mobile phone apps, and computer kiosk), (2) purpose of the interventions (education and support), and (3) key strategies (monitoring and breastfeeding tracking, personalization, online discussion forum, web-based consultation, and breastfeeding station locators). Combining educational activities with web-based personalized support through discussion forums appeared to be the most effective way to improve breastfeeding outcomes and long-term exclusive breastfeeding rates. Monitoring and breastfeeding trackers appeared to be the least effective ways.

Conclusions: This study demonstrated a variety of internet-based e-technologies that professionals can use to promote, educate, and support breastfeeding women. Future internet-based breastfeeding interventions employing e-technologies might consider improving interaction with mothers and personalizing the content of the proposed interventions.

(J Med Internet Res 2020;22(5):e17361) doi:10.2196/17361

KEYWORDS

breastfeeding; mobile app; mobile phone; mHealth; internet; computers; systematic review

Introduction

Breastfeeding provides survival, growth, and health benefits for infants and promotes positive maternal health outcomes [1]. The United Nations Children's Fund and the World Health Organization (WHO) refer to breastfeeding as a cornerstone for child survival, nourishment, and growth and maternal health [2]. However, only 38% of infants aged under 6 months are exclusively breastfed worldwide, which is lower than the 2025 target of 50% [3]. In Australia, 96% of women initiate breastfeeding after birth, but the feeding practices significantly declined to only 15% to 25% at 6 months postnatally [4].

Early breastfeeding cessation has attracted the attention of clinicians, health care providers, and governments to develop

and assess new initiatives. The main interventions to promote and support breastfeeding include home visit counseling [5], peer counseling [6], peer support groups [7], and in-hospital educational intervention [8]. Breastfeeding education with support that starts from early pregnancy and continues to the late postnatal period was demonstrably one of the most effective interventions in long-term breastfeeding behavior [9]. However, providing traditional face-to-face education and support may require trained professionals [10], and it may not be easy for women and their families to attend educational sessions at an inconvenient time or place [11]. Hence, it has been argued that such traditional forms of support and education may not be beneficial or useful for younger generations [12]. Therefore, information and communications technology (ICT) may be suitable in transforming traditional education and support into a free and widely accessible mode of delivery [12,13].

ICTs are defined as digital technology tools and resources used to capture, handle, store, and exchange information via electronic communication [14]. The WHO indicated that ICT improves access to information, which will lead to an improvement in health care services around the world [15]. A study on mobile phone apps used by low-income participants in a public health nutrition program for Women, Infants, and Children in the United States revealed the need for improving and expanding technology in their program [3]. An international board–certified lactation consultant, Heinig [12], proposed that using the internet to deliver breastfeeding interventions is a way forward to promote breastfeeding. From April 2019, almost 4.4 billion people were actively using the internet [16]. Electronic technologies (e-technologies) are regarded as a revolutionary advance for providing health care services. The varied applications of e-technologies demonstrate improved operational efficiency and optimized time and productivity for both patients and health care professionals [10]. E-technologies such as web-based technologies, mobile apps, and computer kiosks use a broader range of ICTs to extend beyond the traditional health care facilities and provide support to geographically distant populations [17-19]. Evidence demonstrates that e-technologies can also deliver personalized web-based interventions that generate a longer-lasting health behavior change [20,21], such as in breastfeeding. Studies have shown a noticeable interest in more web-based options with personalized information, providing support to women on breastfeeding decisions [22-24]. Providing internet-based customizable support to mothers seeking information and help during their breastfeeding experience through e-technologies has the potential to impact any breastfeeding outcomes positively [25,26].

Many systematic reviews have assessed the efficacy of interventions employing e-technologies on maternity care and pregnancy outcomes [27-31]. However, only a few systematic reviews have reported on the effect of interventions employing e-technologies on breastfeeding outcomes. For instance, in a systematic review with 3 electronic interventional studies, using electronic-based interventions had a moderate effect on breastfeeding compared with provider-based interventions (OR 2.2, 95% CI 1.9-2.7; d=0.5 versus OR 1.1, 95% CI 1.0-1.2; d=0.0) [32]. Similarly, Giglia and Binns [33] reviewed references published between 2000 and May 2013 to assess the effect of using the internet on breastfeeding outcomes. The study found that among 1379 studies, only 1 study was eligible for inclusion and demonstrated a positive effect of using the internet on breastfeeding outcomes. The findings of another study on the efficacy of e-technologies in improving breastfeeding outcomes [34] suggested that interventions employing e-technologies are potentially valuable for improving breastfeeding knowledge, initiation, and duration of exclusive breastfeeding. However, there is no current systematic review to identify the different types of contemporary internet-based interventions employing e-technologies and assess their impact on breastfeeding outcomes. This study aimed to identify the features of the current internet-based breastfeeding interventions employing e-technologies and investigate the effects of internet-based e-technologies on breastfeeding outcomes.

Methods

Search Methodology

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were followed in this review [35]. A PRISMA checklist and search terms are available in Multimedia Appendix 1.

Information Sources

A total of 8 electronic databases were searched to identify potential studies: Scopus, Web of Science, the Cochrane Database of Systematic Reviews, ScienceDirect, Google Scholar, the Association for Computing Machinery, SpringerLink, and Institute of Electrical and Electronics Engineers Xplore. The registers of the following trials were also searched to identify any existing relevant trials: Cochrane Central Register of Controlled Trials, WHO International Clinical Trials Registry Platform, International Clinical Trials Registry Platform, and ClinicalTrials.gov.

Types of Studies and Inclusion/Exclusion Criteria

Peer-reviewed studies, including quantitative and quality research, mixed methods, descriptive studies, randomized controlled trials, and quasi-experimental design trials with or without blinding, were included in the review. The research methodology was not limited in any way. Any papers examining interventions employing e-technologies that required internet access and aimed at addressing any breastfeeding outcome were included. Considering that SMS messaging can also be used without internet access, studies that used SMS messaging have been excluded from the review. The inclusion and exclusion criteria are shown in Textboxes 1 and 2.

Textbox 1. Inclusion criteria.

- Topic
  - Papers evaluating any internet-based breastfeeding interventions employing electronic technologies (e-technologies; web-based, computer kiosk, and mobile app)
  - Papers evaluating internet-based breastfeeding interventions employing e-technologies among women
  - Papers explaining, assessing, or reporting any internet-based breastfeeding interventions employing e-technologies with any breastfeeding outcome
- Settings
  - No restriction
- Type of publications
  - Papers published in peer-reviewed journals or peer-reviewed papers from an international academic conference or conference proceeding.
- Language
  - English
- Publication date
  - No restriction

Textbox 2. Exclusion criteria.

- Topic
  - Papers evaluating other breastfeeding interventions employing electronic technologies (e-technologies; eg, phone call, text messaging, and video or phone call)
  - Papers evaluating any internet-based breastfeeding interventions employing e- technologies targeted only at other relevant parties (eg, fathers, clinicians, providers, and health care worker or services)
  - Papers do not report on any breastfeeding outcome but focus on general maternal and child health
- Settings
  - No restriction
- Type of publications
  - Oral presentations, commentaries, policy briefs, and papers that described an app without evaluating its implementation and study protocol.
- Language
  - Not English
- Publication date
  - No restriction

Types of Participants

Participants were healthy pregnant or postnatal women, either primiparous or multiparous, who intended to breastfeed. In addition, studies targeting both parents as study participants were also included if the intervention targeting women was described separately. Women of all ages, ethnicity, occupation, and any socioeconomic status were included.

Search Strategy

The search was conducted on November 2018, whereas the screening stages occurred between January and March 2019. The following main key terms were used in the search strategy: nursing, breast-feed, “breast-feeding,” “breast feeding,” “breast milk” and lactation; and computers, telehealth, computer-mediated and mobile application. A detailed example of the search strategy tailored for Scopus is outlined in Textbox 3. This search strategy was adapted for each of the other databases (Multimedia Appendix 2).

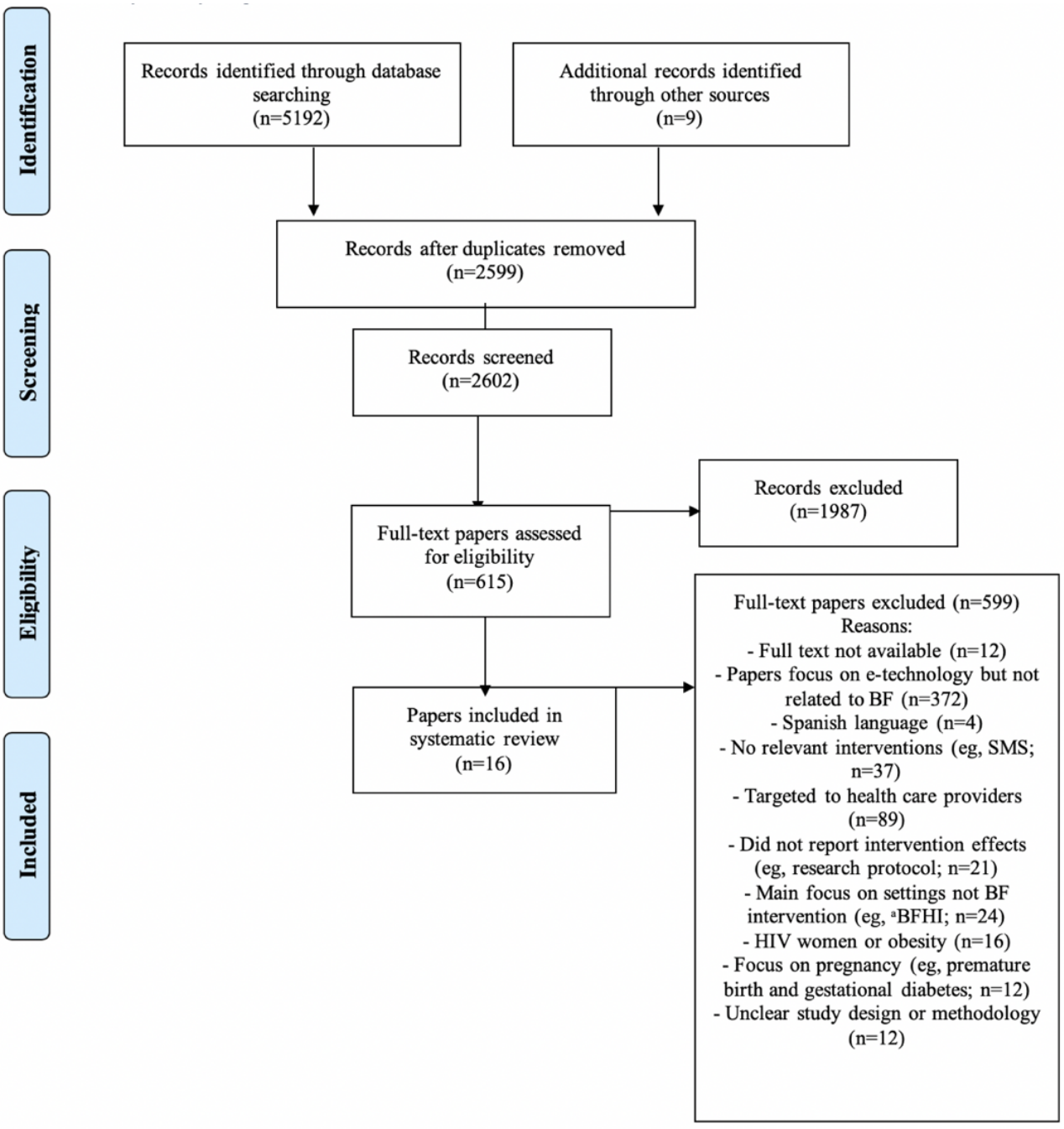
Textbox 3. Search strategy for the Scopus database.

(TITLE-ABS-KEY(Technology OR computer* OR web OR internet OR mobile* OR smartphone OR SMS OR video OR messag* OR application OR intervention OR promotion OR support OR cellphone OR ios OR android OR Cell* OR telephone OR text)) AND (TITLE-ABS-KEY("Mobile health" OR mhealth OR m-health OR e-health OR ehealth OR telemedicine OR Telehealth OR Telelactation OR Health Information Technology OR behavio*)) AND (TITLE-ABS-KEY(Breastfeed* OR Breast-feed OR "Breast feed" OR lacta* OR "nursing mother*" OR mother* OR maternal health OR maternal care OR pregnan* OR antenatal OR post* OR newborn* OR infant* OR child* OR baby OR Exclusive Breast Feeding OR Exclusive Breastfeeding))

Data Extraction

The retrieved studies for the search were transferred to Elsevier's Mendeley Desktop Reference Manager by the first author (AA) to download papers and remove duplicates. The first author screened all titles and abstracts of the papers, and the full texts were reviewed to identify eligible studies. Then, 3 authors (AA, SM, and KW) discussed the suitability of the eligible studies for the final review. Studies that investigated maternal depression, HIV/Hepatitis C virus, smoking, diabetes, alcohol, overweight/obesity, fertility, prematurity, or cesarean section were excluded from the review. A table of the excluded studies with reported reasons is included in Multimedia Appendix 5. A diagram of the systematic review and data extraction is presented in Figure 1.

Figure 1. Flow Diagram of the systematic review - data extraction (BF, BSES). BF: breastfeeding; BHFI: baby friendly hospital initiative.



Data Analysis

A self-adaptive narrative synthesis approach with thematic analysis [36] was used to categorize the purpose of internet-based interventions employing e-technologies and the strategies that were used in those interventions. Meta-analysis could not be processed because of the substantial heterogeneity existing among the studies. However, breastfeeding outcomes are presented in a narrative form.

Search Outcome and Quality Appraisal

The search identified 5201 papers. After excluding duplicates, 2599 papers were screened for eligibility and 615 remained (Figure 1). After examination of the full texts, 17 studies fully met the inclusion criteria. The quality of the studies was appraised by 2 researchers (AA and SM), using the mixed methods appraisal tool [37,38]. After the quality appraisal, 1 study was excluded [39], and a total of 16 studies remained (Figure 1). Any disagreement on quality checks was resolved through discussion between the 3 authors (AA, SM, and KW). Tables showing the quality appraisal process are presented as additional files in Multimedia Appendix 4.

Results

Description of Studies

The included studies involved 4018 participants from 8 different countries: the United States (n=5), Australia (n=3), Finland (n=3), Taiwan (n=1), Canada (n=1), Thailand (n=1), the Philippines (n=1), and Ireland (n=1). The interventions in the included studies were delivered to women prenatally or postnatally, and all reported one or more breastfeeding outcomes (Table 1). Only 3 studies reported using a theoretical framework to inform the design of the intervention: the sustained breastfeeding framework based on predicting and changing behavior theory [40], Bandura's [41] social cognitive theory [42], and Bandura's [43] self-efficacy theory [44]. The description of the included studies is provided in Multimedia Appendix 3.

Characteristics of Internet-Based Electronic Technology Interventions

Characteristics of the internet-based e-technology interventions in the 16 included studies were categorized by the mode of e-technology delivery, purpose of the intervention, and key strategies that were used in those interventions.

Mode of Electronic Technology Delivery

Electronic technology in the included studies involved a range of delivery approaches such as web-based interventions [42,44,46-54], mobile apps [55-57], and computer kiosk [40,45] (Table 2).

Purpose and Key Strategies Used in the Interventions

The purpose and key strategies used in the studies were coded using a thematic analysis. Overall, 2 main categories emerged based on the purpose of the interventions: breastfeeding education and breastfeeding support. On the basis of key strategies that were used in the interventions, 6 categories were identified: (1) monitoring and breastfeeding tracking, (2) personalization, (3) online discussion forum, (4) web-based consultation, and (5) breastfeeding station locator (Table 3).

Breastfeeding Outcome Measures

Breastfeeding intention, knowledge, exclusivity, and duration were the most reported outcome measures in the studies (Table 4).

Table 1. Characteristics of included studies.

Study reference	Setting, study country	Study design	Participants	Duration	Breastfeeding outcome measure	Theoretical framework	Assessment tool
Joshi et al, 2016 [40]	The Regional West Medical Centre in Scottsbluff, United States	Quasi-experimental study	46 women	Antenatal to 6 months postnatal	• Breastfeeding knowledge • Breastfeeding self-efficacy • Breastfeeding intention	Predicting and changing behavior theory	• Breastfeeding knowledge questionnaire • BSES-SF[a] questionnaire • Breastfeeding attrition prediction tool questionnaire
Zhang et al, 2014 [45]	Melrose-Wakefield Hospital in Massachusetts, United States	RCT[b]	15 women	Antenatal to 6 months postnatal	• Breastfeeding knowledge • Breastfeeding intention	NR[c]	• Questionnaire • BSES-SF
Ahmed et al, 2012 [46]	Two Midwestern hospitals, United States	Mixed methods	26 women	Postnatal to 30 days	• Breastfeeding support • Breastfeeding education	NR	• Breastfeeding diary • System usability scale • Perception survey
Ahmed et al, 2016 [42]	3 hospitals in the Midwestern, United States	RCT	141 women	Postnatal to 3 months	• Exclusive breastfeeding • Breastfeeding duration • Breastfeeding intensity	Social cognitive theory	• Paper-based forums • Surveys (online follow-up forums)
Alberdi et al, 2018 [47]	The National Maternity Hospital (Dublin, urban) and Wexford General Hospital (Wexford, rural), Ireland	Feasibility study	100 women	Antenatal to 3 months postnatal	• Breastfeeding duration	NR	• Web-based questionnaires
Geoghegan-Morphet et al, 2014 [48]	Hospital, Canada	Qualitative study	200 women	Postnatal to 6 months	• Breastfeeding education • Breastfeeding outcomes • Breastfeeding support	NR	• Surveys
Giglia et al, 2015 [49]	Hospitals from 4 regional areas, Australia	Longitudinal cohort study	414 women	Postnatal to 12 months	• Breastfeeding support • Breastfeeding initiation • Breastfeeding duration • Exclusive breastfeeding	NR	• Modified tool based on Perth infant feeding study mark II
Grassley et al, 2017 [50]	Website, United States	Cohort study (1 group pre + post)	41 women	Antenatal to 1 month postnatal	• Breastfeeding self-efficacy • Breastfeeding intention • Breastfeeding education	NR	• Questionnaire (pretest and posttest) • BSES-SF
Hannula et al, 2014 [51]	3 public maternity hospitals in the Helsinki Metropolitan area, Finland	Quasi-experimental study	705 women	Antenatal to first week postnatal	• Exclusive breastfeeding • Breastfeeding attitude • Breastfeeding confidence • Breastfeeding coping	NR	• Iowa infant feeding attitude scale • BSES-SF • LATCH[d] assessment tool

Study reference	Setting, study country	Study design	Participants	Duration	Breastfeeding outcome measure	Theoretical framework	Assessment tool
Huang et al, 2007 [52]	Hospital in Taipei, Taiwan	Quasi-experimental study	120 women	Antenatal to 6 weeks postnatal	• Breastfeeding duration • Breastfeeding knowledge • Breastfeeding attitude	NR	• Questionnaire
Newby et al, 2015 [53]	Feeding Queensland Babies, Australia	Prospective cohort study	488 women	Antenatal to 12 months postnatal	• Breastfeeding support • Breastfeeding education	NR	• Web-based questionnaires
Salonen et al, 2008 [44]	Two public university hospitals, Finland	Quasi-experimental study	863 women	Antenatal to hospital discharge	• Exclusive breastfeeding	Self-efficacy theory	• Questionnaire
Salonen et al, 2014 [54]	Hospital, Finland	Quasi-experimental study	760 women	Antenatal to 12 months postnatal	• Exclusive breastfeeding	NR	• Structured questionnaire
Dela Cruz et al, 2017 [55]	The Philippine Human Milk Bank, Philippines	Qualitative study	32 women	NR	• Breastfeeding support • Breastfeeding education	NR	• Web-based questionnaires
Wang et al, 2018 [56]	Website, Thailand	Mixed method	21 women	Postnatal to 4 weeks	• Breastfeeding support • Breastfeeding education	NR	• Surveys • Structured interview
Wheaton et al, 2018 [57]	4 local hospitals at Southwest Victoria, Australia	Prospective cohort study	46 women	Postnatal to 6 months	• Breastfeeding duration • Breastfeeding confidence	NR	• Web-based questionnaires

[a]BSES-SF: breastfeeding self-efficacy scale—short form.

[b]RCT: randomized controlled trial.

[c]NR: not reported.

[d]LATCH: breastfeeding charting system and documentation tool.

Table 2. Description of internet-based electronic technology interventions.

Study reference	Type of technology intervention	Purpose/objectives	Providers
Joshi et al, 2016 [40]	Computer kiosk	To provide breastfeeding knowledge and enhance self-efficacy through an interactive computer kiosk and a bilingual breastfeeding educational program	• Nurses
Zhang et al, 2014 [45]	Computer kiosk	To promote breastfeeding through a virtual lactation consultant on a computer kiosk	• University staff
Ahmed et al, 2012 [46]	Website	To provide breastfeeding education and support through an interactive web-based breastfeeding monitoring system (LACTOR)	• Trained research team • Lactation consultants
Ahmed et al, 2016 [42]	Website	To increase breastfeeding duration, exclusivity, and intensity through a web-based interactive breastfeeding	• Research scientist and a lactation consultant
Alberdi et al, 2018 [47]	Website	To assess breastfeeding initiation and duration using a web-based breastfeeding support	• Lactation specialists
Geoghegan-Morphet et al, 2014 [48]	Website	To provide breastfeeding education with integrated peer and professional support through a web-based breastfeeding support clinic	• Academic Medical Organization of Southwestern Ontario study team
Giglia et al, 2015 [49]	Website	To provide breastfeeding support using a website	• Midwives and/or research staff • Nurses
Grassley et al, 2017 [50]	Website	To provide breastfeeding education, assess breastfeeding intention, and promote breastfeeding self-efficacy using a web-based game-based learning platform	• Researchers' university
Hannula et al, 2014 [51]	Website	To provide an intensified support for breastfeeding using a web-based service	• Midwives
Huang et al, 2007 [52]	Website	To provide breastfeeding education, increase breastfeeding knowledge, and enhance breastfeeding skills using a web-based breastfeeding education program	• Midwives
Newby et al, 2015 [53]	Websites	To provide breastfeeding and education and support using web-based breastfeeding support and evaluate internet sources of infant feeding information	• The researcher
Salonen et al, 2008 [44]	Website	To support breastfeeding, parenting, and infant care through a web-based information website	• Nurses
Salonen et al, 2014 [54]	Website	To support breastfeeding, parenting, and infant care through a web-based information website	• Nurses
Dela Cruz et al, 2017 [55]	Mobile app	To provide breastfeeding education and support using a mobile app	• The researchers
Wang et al, 2018 [56]	Mobile app	To provide breastfeeding education and support through a mobile phone app	• The researchers
Wheaton et al, 2018 [57]	Mobile app	To evaluate breastfeeding duration compared with the general population and describe infant feeding outcomes	• Nurses and/or midwives

Table 3. Classification of the internet-based electronic technologies based on the purpose of the interventions and the key strategies.

Purpose and key strategies	Computer kiosk	Website												Mobile app		
	Joshi et al, 2016 [40]	Zhang et al, 2014 [45]	Ahmed et al, 2016 [42]	Huang et al, 2007 [52]	Hannula et al, 2014 [51]	Salonen et al, 2008 [44]	Salonen et al, 2014 [54]	Geoghegan-Morphet et al, 2014 [48]	Grassley et al, 2017 [50]	Ahmed et al, 2012 [46]	Giglia et al, 2015 [49	Alberdi et al, 2018 [47]	Newby et al, 2015 [53]	Wang et al, 2018 [56]	Dela Cruz et al, 2017 [55]	Wheaton et al, 2018 [57]
Purpose																
Education	✓[a]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	N/A[b]	✓	✓
Support	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Key strategies																
Monitoring and breastfeeding tracking	N/A	✓	✓	N/A	N/A	N/A	N/A	N/A	N/A	✓	N/A	N/A	N/A	✓	N/A	N/A
Personalization	✓	✓	✓	N/A	N/A	N/A	N/A	✓	N/A	✓	✓	N/A	N/A	✓	✓	✓
Online discussion forums	N/A	N/A	N/A	✓	N/A	✓	✓	✓	N/A	N/A	✓	✓	✓	N/A	N/A	N/A
Web-based consultant	N/A	N/A	✓	N/A	N/A	N/A	N/A	✓	N/A	✓	✓	N/A	N/A	N/A	N/A	N/A
Breastfeeding stations locator	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	✓	✓	N/A

[a]✓: applicable.

[b]N/A: not applicable.

Table 4. Reported breastfeeding outcomes based on the types of internet interventions employing electronic technologies.

Outcome measure[a]	Total (N=16), n (%)	Computer kiosk, n	Website, n	Mobile app, n
Breastfeeding initiation	1 (6)	N/A[b]	1	N/A
Exclusive breastfeeding	5 (31)	N/A	5	N/A
Breastfeeding duration	5 (31)	N/A	4	1
Breastfeeding intention	3 (19)	2	1	N/A
Breastfeeding knowledge	3 (19)	2	1	N/A
Breastfeeding attitude	1 (6)	N/A	1	N/A
Breastfeeding confidence	2 (13)	N/A	1	1
Breastfeeding self-efficacy	1 (6)	1	N/A	N/A
Breastfeeding intensity	1 (6)	N/A	1	N/A

[a]Studies with multiple outcomes were counted repeatedly in each electronic technology category.

[b]N/A: not assessed.

Breastfeeding Outcomes With Web-Based Interventions

Of 11 web-based breastfeeding interventions that had a combination of education and support focus, 8 demonstrated improvements in breastfeeding outcomes [42,44,47-49,51,52,54].

For instance, Huang et al [52] evaluated breastfeeding knowledge in a quasi-experimental study (n=120) where there was a significant difference in breastfeeding knowledge level between the intervention and control groups in the posttest results (P<.001), with no differences in the pretest knowledge level. In addition, the exclusive breastfeeding rates were statistically higher at 3 to 5 days and at 2, 4, and 6 weeks in the intervention group (n=60) compared with the control group (n=60; 48.3%, 45%, 31.7%, and 26.7% versus 38.3%, 20%, 20%, and 20%, respectively) [52].

Exclusive breastfeeding during hospital stay has been shown to improve with web-based interventions [44,51,54]. Hannula et al [51] evaluated the effect of a web-based breastfeeding educational support system on exclusive breastfeeding rates in 705 women in Finland. The intervention group (n=431) had access to a website that offered articles, pictures, videos, and an educational game. During hospital stay, the intervention group breastfed exclusively more often than the control group (71% versus 58%; P<.001) and likewise on discharge (76% versus 66%; P=.01). In addition, 2 other quasi-experimental studies conducted in Finland with a convenience sample of 1300 [44,54] also indicated higher exclusive breastfeeding rates in a group of women who used web-based interventions compared with the control group during their stay in the hospital. Salonen et al [44] reported that exclusive breastfeeding was more common in the intervention group (51.0% versus 27.4%; P<.001) and similar findings were also reported by Salonen et al [54] (50.1% versus 30.4%; P<.001).

Providing web-based breastfeeding intervention with an interactive and asynchronous online discussion board was found useful in motivating intervention mothers to continue breastfeeding for longer periods of time [49]. For instance, in a longitudinal cohort study among regional Western Australian women, using online support and discussion forums was associated with higher exclusive breastfeeding rates at 6 months among women who lived in remote areas (n=10; 5.9% versus 0.6%; P=.01) [49].

Access to a website with information on breastfeeding has also been offered to 127 women in a recent study in Ireland along with sending emails up to 6 months postpartum and providing access to a breastfeeding helpline and support group [47]. The study demonstrated a positive impact on breastfeeding duration, as participants from both urban and rural areas acknowledged an increase in breastfeeding duration because of participation in the study, 42.2% and 86.7%, respectively. However, urban women (42% at 6 weeks and 53.3% at 3 months) reported no impact on the length of time they breastfed from participation in the study, compared with rural women (26.7% at 6 weeks and 13.3% at 3 months). This study found that providing additional support postnatally, including exclusive access to the educational website, was more preferred by women in rural areas compared with those in urban areas, with 46.7% and 22%, respectively [47].

Geoghegan-Morphet et al [48,58] designed a web-based breastfeeding support clinic called the Maternal Virtual Infant Nutrition Support clinic to offer evidence-based breastfeeding education combined with peer and professional support. The resource had 4 aspects: (1) text, graphics, and video breastfeeding educational resources; (2) an interactive discussion forum for participants, which is supervised and facilitated by an international board–certified lactation consultant; (3) data collection capacity; and (4) a web-based infant journal for text and photo entries. The study found that 60.8% of mothers were exclusively breastfeeding at 6 months postpartum in the intervention group compared with 28.2% of mothers, similar to statistics from the Canadian 2010-2013 data for Ontario mothers [58].

A US-based pilot study [50] designed the *HealthyMoms* intervention aimed to educate women about breastfeeding using a game-based learning platform (3D Gamelab) on the web. A total of 3 quests on the 3D Gamelab platform were completed by the study participants and focused on three main topics: deciding about breastfeeding, feeding your baby, and getting support. Each one of the quests focused on a specific breastfeeding topic using web-based education activities, such as watching a video, reading a short introduction, or adding posts and responses to the information. Breastfeeding self-efficacy was measured with the 14-item breastfeeding self-efficacy scale—short form (BSES-SF), and breastfeeding intention was assessed using a 1-item measure with 4 categories (eg, *just breastfeed/no formula*, *just formula/no breastfeeding*, *both*, or *unsure*). The majority of mothers (68%) reported a high intention for exclusive breastfeeding; however, a one-way analysis of variance test found no significant differences among the groups in mean BSES-SF scores before (P=.26) or after (P=.68) the intervention [50].

Furthermore, 2 studies evaluated the impact of web-based breastfeeding monitoring systems among breastfeeding mothers [42,46]. Of the systems, one [42] found no significant differences in breastfeeding outcomes between the women in the intervention group (n=49) and the control group (n=57) upon discharge (P=.71). However, the women in the intervention group had better exclusive breastfeeding rates (63%, 63%, and 55%, respectively) at 1, 2, and 3 months, respectively, compared with the control group (40%, 19%, and 19%, respectively). Participants (n=26) in a study by Ahmed and Ouzzani [46] also received postdischarge breastfeeding support through web-based breastfeeding monitoring systems and reported that the system helped them to minimize breastfeeding problems [46].

The Key Strategies Used in the Web-Based Interventions

Monitoring and Breastfeeding Tracking in the Web-Based Interventions

Only 2 studies [42,46] provided web-based breastfeeding diary interventions and used data monitoring strategies to promote breastfeeding.

Providing a web-based breastfeeding diary [42] designed to record breastfeeding data for 30 days had a significant effect on breastfeeding intensity between the intervention and control groups at 3 months (P=.002). Ahmed and Ouzzani [46] evaluated the impact of a web-based breastfeeding monitoring system (LACTOR) designed to record breastfeeding and infant output data for 30 days. The system has 2 modules: (1) the mothers' portal, where mothers can record their daily breastfeeding data and get notifications; and (2) the lactation consultant's portal. The study found that more than 77% of the mothers reported infant feedings ≥8 times per day and reported that the system provided motivation to continue breastfeeding. Although 70% of mothers stated that breastfeeding data entry was not time-consuming, they reported that monitoring did not substitute face-to-face consultation with a lactation consultant.

Personalization in the Web-Based Interventions

Delivering personalized support was cited in four studies [42,46,48,49]. In all, two web-based breastfeeding monitoring systems [42,46], asked participating mothers to enter their breastfeeding data daily where both systems are capable of generating an automatic feedback sent via notifications with tailored interventions depending on the entered breastfeeding problem. These systems are capable of detecting patterns from responses of mothers and recognizing different breastfeeding problems. The system also provided motivation for mothers in the intervention group by sending a positive notification when a mother breastfed 8 to 10 times per day [42]. A website breastfeeding intervention in Australia enabled the intervention group (n=207) to post on discussion forums, start a new email conversation with others, and use a webcam to contact lactation consultants and receive a tailored response [49]. A more personalized intervention was proposed in a study by Geoghegan-Morphet et al [48], where a lactation consultant offered live help sessions through the web-based clinic and real-time response; however, these help sessions were not utilized by any of the participating women.

Online Discussion Forums in the Web-Based Interventions

Providing a web-based breastfeeding intervention with an interactive and asynchronous online discussion board was found to be effective in increasing rates of exclusive breastfeeding [52] and useful in motivating mothers in the intervention group to continue breastfeeding for longer [49]. Discussion forums were also used by mothers to contact a registered nurse/midwife to ask questions and receive credible information [44,54]. However, no relationship was discovered between accessing online chatrooms/discussion forums for breastfeeding support and providing breast milk for infants [53]. Remarkably, the breastfeeding helpline in an Irish cohort [47] was accessed by 30.6% of urban women and only 12.5% of rural women.

The web-based breastfeeding support clinic developed in a study in Canada [48] improved access to specialized professional breastfeeding support and offered communication and engagement through discussion forums. The intervention group received full access with peer and professional support in the discussion forum, and the control group received full access except for the interactive discussion forum [48].

Web-Based Lactation Consultant in the Web-Based Interventions

In an internet-based intervention study by Giglia et al [49], which was effective in long-term breastfeeding outcomes, women in the intervention group were provided with access to a certified lactation consultant through web-based posts or using a webcam with any of their concerns or questions regarding breastfeeding. Although women in this study reported several breastfeeding issues at each time point, none of them contacted a lactation consultant through webcam services.

Similarly, a web-based breastfeeding support clinic [48] offered support and educational resources to mothers facilitated by a lactation consultant where they could monitor, facilitate, and encourage discussions on the discussion forum. They could also post information and resources, set questions for additional details, answer questions, and offer suggestions. However, the post topics from the intervention group tended to be on more subjective (eg, lifestyle based) rather than technical topics (eg, breastfeeding problems) [48]. In 2 studies [42,46], the breastfeeding data of mothers were continuously monitored on the web by lactation consultants. Early detection and resolution of breastfeeding problems by lactation consultants presented improvements to breastfeeding continuation.

Breastfeeding Outcomes With Using Mobile Apps

A total of 3 studies reported on the usability of mobile apps without examining the effectiveness of the app on breastfeeding outcomes [55-57]. One Australian study evaluated the usability of a smartphone breastfeeding app, *Breastfeeding Solutions*, among rural Australian breastfeeding women [57]. The app was an interactive guide to resolve breastfeeding problems, provide functions for searching for problem solutions, and deliver timely information for mothers. The results of the study demonstrated its usability by a longer duration of breastfeeding compared with general statistics from southwest Victoria [57].

A total of 2 other studies with small sample sizes evaluated the usability and usefulness of the mobile phone apps *MoomMae* [56] *and Milktrack* [55] designed to provide support to breastfeeding women. The *MoomMae* [56] intervention aims to support mothers in feeling more comfortable breastfeeding in public and efficiently track their feeding and pumping logs.

Key Strategies Used in Mobile Device Apps

Monitoring and Breastfeeding Tracking in Mobile Apps

Only one breastfeeding mobile app intervention [56] offered breastfeeding tracking features to enter pumping/feeding data and save the history of all feeding records.

Personalization in Mobile Apps

Participants enrolled in 3 studies [55-57] received a personalized breastfeeding intervention through mobile apps. The interventions were designed to enhance the breastfeeding experience by providing feeding and pumping volume control [56], offering a platform to donate human milk [55], and providing an interactive guide to solving the breastfeeding problems of mothers [57].

Breastfeeding Station Locator in Mobile Apps

In all, 2 studies in Thailand [56] and the Philippines [55] developed mobile apps with breastfeeding station locators with an embedded strategy that can locate nearby places accessible for breastfeeding using the user's GPS location. In both studies, women perceived that locating places for breastfeeding through the app was easier, but the *feeding room* feature still had some negative feedback because of the limited number of feeding rooms exhibited in the app [56].

Breastfeeding Outcomes Using a Computer Kiosk

A total of 2 studies used computer kiosk interventions [40,45]. The main purpose of using computer kiosks was to provide breastfeeding education and support by incorporating several educational modules such as basics of breastfeeding, benefits of breastfeeding, and coping with breastfeeding [40,45]. A pilot evaluation study [45] showed significantly greater breastfeeding knowledge in the intervention group (n=7) after they interacted with the prenatal module (P<.05) in comparison with the control group (n=8). This study showed significantly greater intention to exclusively breastfeed after women interacted with the prenatal module (P<.05) and improvement in breastfeeding confidence compared with the control group [45]. Similarly, Joshi et al [40] reported a significant improvement in breastfeeding knowledge scores (P=.03) only at week 6 of the follow-up between the control (mean 23.2, SD 3.7) and intervention (mean 25.3, SD 2.6) groups. However, no significant differences were perceived in the average change in knowledge scores between the control and intervention groups at any other follow-up time points. This study also reported a gradual increase in the breastfeeding self-efficacy scores until week 6, followed by a decrease in self-efficacy scores at 3 months (P=.46) and 6 months (P=.54). Moreover, the intervention group reported significantly higher intention to breastfeed (P=.049), and the results indicated a significant improvement in breastfeeding intention scores over a 6-month period with all study participants (P<.05).

Key Strategies Used in Computer Kiosks

Monitoring and Breastfeeding Tracking With Computer Kiosks

Only the computer kiosks in the Zhang et al [45] study had a longitudinal breastfeeding tracking feature, which enabled the system to examine breastfeeding records and monitor breastfeeding practices to ensure optimal infant growth.

Personalization With Computer Kiosk

Zhang et al [45] developed a virtual lactation consultant on a computer kiosk to interact with women in the intervention group. The intervention included motivational interviewing techniques to motivate and social cognitive techniques to reinforce positive behaviors. The intervention enabled adaptive interaction, where the interactive component was modifiable based on a mother's own progress and her previous interactions with the kiosk. Joshi et al [40] designed an interactive kiosk to present breastfeeding information and messages adjustable depending on the psychosocial elements, including self-efficacy, influence of attitude, expectancies, personal norms, and social effect.

Discussion

Principal Findings

This study presents a review of internet-based breastfeeding interventions employing e-technologies; investigates their purpose, mode of delivery, and key strategies; and systematically describes their effectiveness on breastfeeding outcomes. The main purpose of the reviewed studies focused on education and support in 3 modes of delivery: web-based interventions, mobile apps, and computer kiosks. Among the 5 key strategies, personalization and web-based discussion were the most common strategies used in the interventions. This review provides evidence that internet-based interventions employing e-technologies that provide a combination of early interactive antenatal breastfeeding education with postnatal web-based discussion support can improve breastfeeding outcomes in hospital stay [44,47,48,51,52,54] and exclusive breastfeeding rates up to 6 months [49]. These results are supported by similar face-to-face interventional studies where there is a combination of antenatal education and postnatal support [59-61].

Interventions with personalized feedback and tailored information to mothers through their support systems were found to be motivational and led to positive breastfeeding outcomes. These effective strategies point toward a need for informed, highly interactive, and tailored-designed breastfeeding interventions employing e-technologies. Consistent with other effective clinical interventions, conducting multimodal, multiphased, and interactive interventions are successful in breastfeeding practices [60,62,63]. Web-based support could be offered informally, such as peer communication through online discussion forums, or formally, such as web-based lactation consultation. The interactivity, connectivity, and two-way communication provided between mothers and lactation consultants in breastfeeding interventions may expand the opportunities for educating and engaging mothers and thus improve breastfeeding outcomes. Providing access to timely and qualified lactation consultants in similar interventions has been found to be particularly attractive to mothers experiencing some breastfeeding complications [64]. In a Cochrane review of support interventions, the findings demonstrated that women who receive any form of support are less likely to stop exclusive breastfeeding before 5 months postpartum [65].

In this review, interventions that provided monitoring and breastfeeding tracking appeared to be least effective in improving breastfeeding outcomes, which could be because of a lack of meaningful educational and supportive interactions for different breastfeeding challenges. It is important to note that women in the included studies had higher prenatal breastfeeding intention rates, which could be a confounding factor in the effectiveness of the interventions [40,45,50]. Breastfeeding women are very likely to seek out extensive evidence-based information on breastfeeding information from reliable sources. The acquired breastfeeding knowledge is presumed to have positive influences on their feeding decisions [66].

In terms of the mode of delivery, web-based platforms were the dominant and more effective mode of delivering breastfeeding

internet-based interventions employing e-technologies. Several studies used web-based interventions, and the majority of them reported an increase in exclusive breastfeeding after the intervention and overall positive outcomes in their results [42,44,47,49,51,52,54]. This finding may be explained by the ability of web-based interventions to sustain their effects in terms of providing early education and continued support to participants during the intervention from a variety of care providers (eg, peers and lactation consultants) in a range of settings (eg, hospital and home). Another reason could be related to the reliability and credibility of the governmental web-based platforms [67]. Women need to make informed decisions regarding the health and well-being of their babies. Having a trusted resource would empower women to make their own choices in breastfeeding. However, many studies have reported the lack of a regulatory system to assist end users in identifying the best available web-based e-technologies [68-70].

Although the delivery of breastfeeding interventions through mobile apps is encouraging because of their relative simplicity and continuous availability [71], the mobile apps in this review focused on supporting women to access milk banks or express their breastmilk among small groups of women. Furthermore, the studies mainly aimed at assessing the usability of the apps rather than examining their effectiveness without any theoretical framework. Although a significant absence in reporting a theoretical framework was observed in the majority of the included studies, none of the interventions with mobile apps reported using any type of behavior change theoretical framework. Using theory to inform intervention development or evaluation can play a key role in breastfeeding interventions as a strong predictor of breastfeeding behavior [72].

Finally, computer kiosk interventions seemed to be effective in enhancing women's knowledge and short-term confidence level. Creating an interactive platform coupled with continuing, accessible support can aid mothers in having specific information to suit their personal needs [45,73]. The use of conversational computers could provide affordability and portability, as observed in other health-related interventions [74,75]. However, for long-term effects, there is a need to incorporate personalized approaches and professional support into interventions to create a practical response to the personal needs of mothers. In addition, sustainability remains an issue because these types of interventions could be outdated [76], and findings cannot be generalized to other settings.

A lack of detailed reporting of rigorous intervention development, evaluation, and the content of the implementation intervention was also observed, which makes it difficult to determine the impact of interventions on breastfeeding outcomes [77]. Future studies could design intervention implementation and evaluation guided by a theory toward a comprehensive intervention development process and for a better chance of effective breastfeeding intervention.

Limitations

This systematic review has as a key strength that it followed the PRISMA guidelines [35], an established methodology for ensuring transparent reporting of systematic reviews. However, there are several limitations to this study; for example, it was difficult to generalize the findings observed from several studies because of their small sample size, different timing of studies, and lack of clarity on start and end points. The substantial heterogeneity of the intervention's outcome measures and lack of clear definitions of breastfeeding outcomes in all the included studies added more complexity and prevented meta-analysis.

Another limitation of the study is that most of the included interventions were delivered and published in developed and high-income countries with good quality health care systems [78], which may not be generalizable to women in developing countries or those with diverse socioeconomic status and cultural background [29]. Setting up breastfeeding intervention in regions with inadequate health care providers and resources should be prioritized to improve breastfeeding practices [79]. Furthermore, the approaches of the interventions were focused on changing the behavior of individuals and lacked some key aspects related to cultural differences that may influence the breastfeeding practices of women and could impede them from achieving the desired outcome. This is particularly essential as the breastfeeding decisions women make differ according to whether their own culture is supportive of breastfeeding or not and on the cultural context in which they live. Thus, findings need further investigation regarding their transferability. Applicability needs to be considered when delivering breastfeeding interventions to women in different social milieu.

Conclusions

The findings of the study demonstrate that internet-based e-technologies are transforming the access and delivery of breastfeeding interventions and that they have a considerable potential to assist breastfeeding mothers when seeking support and advice about breastfeeding. The results show that web-based interventions that provide a combination of education and ongoing support are the best models of interventions employing e-technologies to support long-term breastfeeding outcomes. In addition, the review reveals that the two dominant effective strategies are personalization, and online discussion forums form credible sources. Further studies need to explore the usability and effectiveness of interventions employing e-technologies that have theory-based systems designs that could incorporate encouragement and discussion opportunities from credible social and professional sources. The sociocultural needs of women also need to be integrated into these technologies to provide culturally tailored breastfeeding support.

Acknowledgments

PhD scholarship support from the Ministry of Education, Kingdom of Saudi Arabia, to AA is gratefully acknowledged.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Preferred Reporting Items for Systematic Reviews and Meta-Analyses 2009 checklist.
[DOCX File , 64 KB - jmir_v22i5e17361_app1.docx]

Multimedia Appendix 2
Database search strategies and keywords.
[DOCX File , 22 KB - jmir_v22i5e17361_app2.docx]

Multimedia Appendix 3
Table of the characteristics of included experimental studies.
[DOCX File , 40 KB - jmir_v22i5e17361_app3.docx]

Multimedia Appendix 4
Mixed methods appraisal tool, version 2018.
[DOCX File , 81 KB - jmir_v22i5e17361_app4.docx]

Multimedia Appendix 5
Excluded studies.
[DOCX File , 27 KB - jmir_v22i5e17361_app5.docx]

References

1. Horta BL, Bahl R, Martines JC, Victora CG. Evidence on the Long-Term Effects of Breastfeeding: Systematic Review and Meta-Analyses. Geneva, Switzerland: World Health Organization; 2007.
2. World Health Organization. Protecting, Promoting and Supporting Breastfeeding in Facilities Providing Maternity and Newborn Services. Geneva, Switzerland: World Health Oragnization; 2017.
3. Weber SJ, Dawson D, Greene H, Hull PC. Mobile Phone Apps for Low-Income Participants in a Public Health Nutrition Program for Women, Infants, and Children (WIC): Review and Analysis of Features. JMIR Mhealth Uhealth 2018 Nov 19;6(11):e12261 [FREE Full text] [doi: 10.2196/12261] [Medline: 30455172]
4. Australian Government: Department of Health. 2018. Australian National Breastfeeding Strategy: 2019 and Beyond URL: https://consultations.health.gov.au/population-health-and-sport-division/breastfeeding/ [accessed 2020-05-13]
5. Morrow AL, Guerrero ML, Shults J, Calva JJ, Lutter C, Bravo J, et al. Efficacy of home-based peer counselling to promote exclusive breastfeeding: a randomised controlled trial. Lancet 1999 Apr 10;353(9160):1226-1231. [doi: 10.1016/S0140-6736(98)08037-4] [Medline: 10217083]
6. Haider R, Ashworth A, Kabir I, Huttly SR. Effect of community-based peer counsellors on exclusive breastfeeding practices in Dhaka, Bangladesh: a randomised controlled trial [see commments]. Lancet 2000 Nov 11;356(9242):1643-1647. [doi: 10.1016/s0140-6736(00)03159-7] [Medline: 11089824]
7. Kruske S, Schmied V, Cook M. The 'earlybird' gets the breastmilk: findings from an evaluation of combined professional and peer support groups to improve breastfeeding duration in the first eight weeks after birth. Matern Child Nutr 2007 Apr;3(2):108-119. [doi: 10.1111/j.1740-8709.2007.00078.x] [Medline: 17355443]
8. Labarere J, Bellin V, Fourny M, Gagnaire J, Francois P, Pons J. Assessment of a structured in-hospital educational intervention addressing breastfeeding: a prospective randomised open trial. BJOG 2003 Sep;110(9):847-852 [FREE Full text] [Medline: 14511968]
9. Meedya S, Fernandez R, Fahy K. Effect of educational and support interventions to increase long-term breastfeeding rates in primiparous women: a systematic review protocol. JBI Database System Rev Implement Rep 2017 Mar;15(3):594-602. [doi: 10.11124/JBISRIR-2016-003132] [Medline: 28267020]
10. Prado C, Silva IA, Soares AV, Aragaki IM, Shimoda GT, Zaniboni VF, et al. [Nursing contributions to the development of the Brazilian telehealth lactation support program]. Rev Esc Enferm USP 2013 Aug;47(4):990-996 [FREE Full text] [doi: 10.1590/S0080-623420130000400031] [Medline: 24310701]
11. Romano AM. A changing landscape: implications of pregnant women's internet use for childbirth educators. J Perinat Educ 2007;16(4):18-24 [FREE Full text] [doi: 10.1624/105812407X244903] [Medline: 18769519]
12. Heinig MJ. Breastfeeding promotion for generations X and Y: why the old ways won't work. J Hum Lact 2009 Aug;25(3):263-265. [doi: 10.1177/0890334409341450] [Medline: 19666853]

13. Meedya S, Win K, Yeatman H, Fahy K, Walton K, Burgess L, et al. Developing and testing a mobile application for breastfeeding support: the milky way application. Women Birth 2020 Feb 17:- epub ahead of print. [doi: 10.1016/j.wombi.2020.02.006] [Medline: 32081557]
14. Gagnon M, Légaré F, Labrecque M, Frémont P, Pluye P, Gagnon J, et al. Interventions for promoting information and communication technologies adoption in healthcare professionals. Cochrane Database Syst Rev 2009 Jan 21(1):CD006093 [FREE Full text] [doi: 10.1002/14651858.CD006093.pub2] [Medline: 19160265]
15. World Health Organization. 2006. The World Health Report 2006 - Working Together for Health URL: https://www.who.int/whr/2006/en/ [accessed 2020-05-06]
16. Statista. 2019. Digital Population in South Africa as of January 2020 URL: https://www.statista.com/statistics/685134/south-africa-digital-population/ [accessed 2020-05-06]
17. Bensley RJ, Hovis A, Horton KD, Loyo JJ, Bensley KM, Phillips D, et al. Accessibility and preferred use of online web applications among WIC participants with internet access. J Nutr Educ Behav 2014;46(3 Suppl):S87-S92. [doi: 10.1016/j.jneb.2014.02.007] [Medline: 24810002]
18. Holm K, Brødsgaard A, Zachariassen G, Smith AC, Clemensen J. Participatory design methods for the development of a clinical telehealth service for neonatal homecare. SAGE Open Med 2017;5:2050312117731252 [FREE Full text] [doi: 10.1177/2050312117731252] [Medline: 28975028]
19. Almathami HKY, Win KT, Vlahu-Gjorgievska E. Barriers and Facilitators That Influence Telemedicine-Based, Real-Time, Online Consultation at Patients' Homes: Systematic Literature Review. J Med Internet Res 2020 Feb 20;22(2):e16407 [FREE Full text] [doi: 10.2196/16407] [Medline: 32130131]
20. Rosa C, Campbell AN, Miele GM, Brunner M, Winstanley EL. Using e-technologies in clinical trials. Contemp Clin Trials 2015 Nov;45(Pt A):41-54 [FREE Full text] [doi: 10.1016/j.cct.2015.07.007] [Medline: 26176884]
21. Noar S, Harrington NG, van Stee SK, Aldrich RS. Tailored health communication to change lifestyle behaviors. Am J Lifestyle Med 2010 Nov 19;5(2):112-122. [doi: 10.1177/1559827610387255]
22. Asiodu IV, Waters CM, Dailey DE, Lee KA, Lyndon A. Breastfeeding and use of social media among first-time African American mothers. J Obstet Gynecol Neonatal Nurs 2015;44(2):268-278 [FREE Full text] [doi: 10.1111/1552-6909.12552] [Medline: 25712127]
23. Regan S, Brown A. Experiences of online breastfeeding support: support and reassurance versus judgement and misinformation. Matern Child Nutr 2019 Oct;15(4):e12874. [doi: 10.1111/mcn.12874] [Medline: 31299699]
24. The California WIC Association. 2016. MyWic: Updating Wic for A New Generation URL: http://www.calwic.org/storage/documents/Engaging_Families/MyWic_Millennial_report.pdf [accessed 2020-05-06]
25. Friedman LB, Silva M, Smith K. A focus group study observing maternal intention to use a WIC education app. Am J Health Behav 2018 Nov 1;42(6):110-123. [doi: 10.5993/AJHB.42.6.11] [Medline: 30158006]
26. Burman A, Papazafeiropoulou A, Barnett J. The Use of the Internet to Support Breastfeeding: An Exploratory Study of Breastfeeding Supporters' Experiences. In: Proceedings of the IADIS International Conference on e-Society. 2011 Presented at: IADIS'11; March 10-13, 2011; Avila, Spain. [doi: 10.7748/phc.21.9.13.s11]
27. Watterson JL, Walsh J, Madeka I. Using mhealth to improve usage of antenatal care, postnatal care, and immunization: a systematic review of the literature. Biomed Res Int 2015;2015:153402 [FREE Full text] [doi: 10.1155/2015/153402] [Medline: 26380263]
28. Griffiths SE, Parsons J, Naughton F, Fulton EA, Tombor I, Brown KE. Are digital interventions for smoking cessation in pregnancy effective? A systematic review and meta-analysis. Health Psychol Rev 2018 Dec;12(4):333-356. [doi: 10.1080/17437199.2018.1488602] [Medline: 29912621]
29. Colaci D, Chaudhri S, Vasan A. mHealth interventions in low-income countries to address maternal health: a systematic review. Ann Glob Health 2016;82(5):922-935 [FREE Full text] [doi: 10.1016/j.aogh.2016.09.001] [Medline: 28283147]
30. Sherifali D, Nerenberg KA, Wilson S, Semeniuk K, Ali MU, Redman LM, et al. The effectiveness of ehealth technologies on weight management in pregnant and postpartum women: systematic review and meta-analysis. J Med Internet Res 2017 Oct 13;19(10):e337 [FREE Full text] [doi: 10.2196/jmir.8006] [Medline: 29030327]
31. O'Brien OA, McCarthy M, Gibney ER, McAuliffe FM. Technology-supported dietary and lifestyle interventions in healthy pregnant women: a systematic review. Eur J Clin Nutr 2014 Jul;68(7):760-766. [doi: 10.1038/ejcn.2014.59] [Medline: 24781682]
32. Pate B. A systematic review of the effectiveness of breastfeeding intervention delivery methods. J Obstet Gynecol Neonatal Nurs 2009;38(6):642-653. [doi: 10.1111/j.1552-6909.2009.01068.x] [Medline: 19930278]
33. Giglia R, Binns C. The effectiveness of the internet in improving breastfeeding outcomes: a systematic review. J Hum Lact 2014 May;30(2):156-160. [doi: 10.1177/0890334414527165] [Medline: 24646682]
34. Lau Y, Htun TP, Tam WS, Klainin-Yobas P. Efficacy of e-technologies in improving breastfeeding outcomes among perinatal women: a meta-analysis. Matern Child Nutr 2016 Jul;12(3):381-401. [doi: 10.1111/mcn.12202] [Medline: 26194599]
35. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol 2009 Oct;62(10):1006-1012. [doi: 10.1016/j.jclinepi.2009.06.005] [Medline: 19631508]

36. Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis. Int J Qual Methods 2017 Oct 2;16(1):160940691773384. [doi: 10.1177/1609406917733847]
37. Hong Q, Pluye P, Fàbregues S, Bartlett G, Boardman F, Cargo M, et al. Mixed Methods Appraisal Tool. 2018. URL: http://mixedmethodsappraisaltoolpublic.pbworks.com/w/file/fetch/127916259/MMAT_2018_criteria-manual_2018-08-01_ENG.pdf%0Ahttp://mixedmethodsappraisaltoolpublic.pbworks.com/ [accessed 2020-05-06]
38. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. The mixed methods appraisal tool (MMAT) version 2018 for information professionals and researchers. Educ Inf IOS Press 2018 Dec 18;34(4):285-291. [doi: 10.3233/EFI-180221]
39. Sinclair M, Lazenbatt A, Salmon S, Calvert J. Telesupport in breastfeeding. Comput Educ 2000 Apr;34(3-4):341-343. [doi: 10.1016/s0360-1315(99)00056-1]
40. Joshi A, Amadi C, Meza J, Aguire T, Wilhelm S. Evaluation of a computer-based bilingual breastfeeding educational program on breastfeeding knowledge, self-efficacy and intent to breastfeed among rural Hispanic women. Int J Med Inform 2016 Jul;91:10-19. [doi: 10.1016/j.ijmedinf.2016.04.001] [Medline: 27185505]
41. Bandura A. Social cognitive theory: an agentic perspective. Annu Rev Psychol 2001;52:1-26. [doi: 10.1146/annurev.psych.52.1.1] [Medline: 11148297]
42. Ahmed AH, Roumani AM, Szucs K, Zhang L, King D. The effect of interactive web-based monitoring on breastfeeding exclusivity, intensity, and duration in healthy, term infants after hospital discharge. J Obstet Gynecol Neonatal Nurs 2016;45(2):143-154 [FREE Full text] [doi: 10.1016/j.jogn.2015.12.001] [Medline: 26779838]
43. Bandura A. Self-Efficacy: The Exercise of Control. New York, USA: Worth Publishers; 1997.
44. Salonen AH, Kaunonen M, Astedt-Kurki P, Järvenpää AH, Tarkka M. Development of an internet-based intervention for parents of infants. J Adv Nurs 2008 Oct;64(1):60-72. [doi: 10.1111/j.1365-2648.2008.04759.x] [Medline: 18808593]
45. Zhang Z, Bickmore T, Mainello K, Mueller M, Foley M, Jenkins L, et al. Maintaining Continuity in Longitudinal, Multi-method Health Interventions Using Virtual Agents: The Case of Breastfeeding Promotion. In: Proceedings of the International Conference on Intelligent Virtual Agents. 2014 Presented at: IVA'14; August 26-29, 2014; Boston, MA, USA p. 504-513. [doi: 10.1007/978-3-319-09767-1_61]
46. Ahmed AH, Ouzzani M. Interactive web-based breastfeeding monitoring: feasibility, usability, and acceptability. J Hum Lact 2012 Nov;28(4):468-475. [doi: 10.1177/0890334412451869] [Medline: 22864059]
47. Alberdi G, O'Sullivan EJ, Scully H, Kelly N, Kincaid R, Murtagh R, et al. A feasibility study of a multidimensional breastfeeding-support intervention in Ireland. Midwifery 2018 Mar;58:86-92. [doi: 10.1016/j.midw.2017.12.018] [Medline: 29324318]
48. Geoghegan-Morphet N, Yuen D, Rai E, Angelini M, Christmas M, da Silva O. Development and implementation of a novel online breastfeeding support resource: the maternal virtual infant nutrition support clinic. Breastfeed Med 2014 Dec;9(10):520-523. [doi: 10.1089/bfm.2014.0051] [Medline: 25244066]
49. Giglia R, Cox K, Zhao Y, Binns CW. Exclusive breastfeeding increased by an internet intervention. Breastfeed Med 2015;10(1):20-25. [doi: 10.1089/bfm.2014.0093] [Medline: 25358119]
50. Grassley JS, Connor KC, Bond L. Game-based online antenatal breastfeeding education: a pilot. Appl Nurs Res 2017 Feb;33:93-95 [FREE Full text] [doi: 10.1016/j.apnr.2016.10.011] [Medline: 28096030]
51. Hannula LS, Kaunonen ME, Puukka PJ. A study to promote breast feeding in the Helsinki metropolitan area in Finland. Midwifery 2014 Jun;30(6):696-704. [doi: 10.1016/j.midw.2013.10.005] [Medline: 24210842]
52. Huang MZ, Kuo S, Avery MD, Chen W, Lin K, Gau M. Evaluating effects of a prenatal web-based breastfeeding education programme in Taiwan. J Clin Nurs 2007 Aug;16(8):1571-1579. [doi: 10.1111/j.1365-2702.2006.01843.x] [Medline: 17655546]
53. Newby R, Brodribb W, Ware RS, Davies PS. Internet use by first-time mothers for infant feeding support. J Hum Lact 2015 Aug;31(3):416-424. [doi: 10.1177/0890334415584319] [Medline: 25925637]
54. Salonen AH, Pridham KF, Brown RL, Kaunonen M. Impact of an internet-based intervention on Finnish mothers' perceptions of parenting satisfaction, infant centrality and depressive symptoms during the postpartum year. Midwifery 2014 Jan;30(1):112-122. [doi: 10.1016/j.midw.2013.02.009] [Medline: 23623471]
55. Dela CD, Mendoza D. Milktrack: Design and Development of Mobile Application and Logistics System in Empowering Breastfeeding Practice in the Philippines. In: Proceedings of the 2017 IEEE Region 10 Conference. 2017 Presented at: TENCON'17; November 5-8, 2017; Penang, Malaysia. [doi: 10.1109/tencon.2017.8228234]
56. Wang C, Chaovalit P, Pongnumkul S. A breastfeed-promoting mobile app intervention: usability and usefulness study. JMIR Mhealth Uhealth 2018 Jan 26;6(1):e27 [FREE Full text] [doi: 10.2196/mhealth.8337] [Medline: 29374000]
57. Wheaton N, Lenehan J, Amir L. Evaluation of a breastfeeding app in rural Australia: prospective cohort study. J Hum Lact 2018 Nov;34(4):711-720. [doi: 10.1177/0890334418794181] [Medline: 30235043]
58. Silva OD, Rai E, Angelini M, Christmas M, Geoghegan-Morphet N. Maternal virtual infant nutrition support (mavins) clinic. Paed Child Healt 2015;5:e102-e120. [doi: 10.1093/pch/20.5.e102a]

59. Meedya S, Fernandez R, Fahy K. Effect of educational and support interventions on long-term breastfeeding rates in primiparous women: a systematic review and meta-analysis. JBI Database System Rev Implement Rep 2017 Sep;15(9):2307-2332. [doi: 10.11124/JBISRIR-2016-002955] [Medline: 28902698]
60. Meedya S, Fahy K, Parratt J. The Milky Way educational and support programme: structure, content and strategies. Women Birth 2016 Aug;29(4):388-393. [doi: 10.1016/j.wombi.2016.01.006] [Medline: 26868045]
61. Kaunonen M, Hannula L, Tarkka M. A systematic review of peer support interventions for breastfeeding. J Clin Nurs 2012 Jul;21(13-14):1943-1954. [doi: 10.1111/j.1365-2702.2012.04071.x] [Medline: 22672457]
62. Su LL, Chong YS, Chan YH, Chan YS, Fok D, Tun KT, et al. Antenatal education and postnatal support strategies for improving rates of exclusive breast feeding: randomised controlled trial. Br Med J 2007 Sep 22;335(7620):596 [FREE Full text] [doi: 10.1136/bmj.39279.656343.55] [Medline: 17670909]
63. Meedya S, Fahy K, Yoxall J, Parratt J. Increasing breastfeeding rates to six months among nulliparous women: a quasi-experimental study. Midwifery 2014 Mar;30(3):e137-e144. [doi: 10.1016/j.midw.2013.12.010] [Medline: 24485838]
64. Mattar CN, Chong Y, Chan Y, Chew A, Tan P, Chan Y, et al. Simple antenatal preparation to improve breastfeeding practice: a randomized controlled trial. Obstet Gynecol 2007 Jan;109(1):73-80. [doi: 10.1097/01.AOG.0000249613.15466.26] [Medline: 17197590]
65. Britton C, McCormick F, Renfrew M, Wade A, King S. Support for breastfeeding mothers. Cochrane Database Syst Rev 2007 Jan 24(1):CD001141. [doi: 10.1002/14651858.CD001141.pub3] [Medline: 17253455]
66. Radzyminski S, Callister LC. Mother's beliefs, attitudes, and decision making related to infant feeding choices. J Perinat Educ 2016;25(1):18-28 [FREE Full text] [doi: 10.1891/1058-1243.25.1.18] [Medline: 26848247]
67. Meedya S, Sheikh M, Win K, Halcomb E. Evaluation of Breastfeeding Mobile Health Applications Based on the Persuasive System Design Model. In: Proceedings of the International Conference on Persuasive Technology. 2019 Presented at: PERSUASIVE'19; April 9-11, 2019; Limassol, Cyprus p. 201. [doi: 10.1007/978-3-030-17287-9_16]
68. Stevens DJ, Jackson JA, Howes N, Morgan J. Obesity surgery smartphone apps: a review. Obes Surg 2014 Jan;24(1):32-36. [doi: 10.1007/s11695-013-1010-3] [Medline: 23749609]
69. Rosser BA, Eccleston C. Smartphone applications for pain management. J Telemed Telecare 2011;17(6):308-312. [doi: 10.1258/jtt.2011.101102] [Medline: 21844177]
70. Carter T, O'Neill S, Johns N, Brady RR. Contemporary vascular smartphone medical applications. Ann Vasc Surg 2013 Aug;27(6):804-809. [doi: 10.1016/j.avsg.2012.10.013] [Medline: 23535521]
71. Colaci D, Chaudhri S, Vasan A. mHealth interventions in low-income countries to address maternal health: a systematic review. Ann Glob Health 2016;82(5):922-935 [FREE Full text] [doi: 10.1016/j.aogh.2016.09.001] [Medline: 28283147]
72. Sutton S. Using social cognition models to develop health behaviour interventions: problems and assumptions. In: Rutter D, Quine L, editors. Changing Health Behaviour: Intervention and Research with Social Cognition Models. Berkshire, UK: Open University Press; 2002:193-208.
73. Akesson KM, Saveman B, Nilsson G. Health care consumers' experiences of information communication technology-a summary of literature. Int J Med Inform 2007 Sep;76(9):633-645. [doi: 10.1016/j.ijmedinf.2006.07.001] [Medline: 16931133]
74. 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]
75. Bickmore T, Gruber A. Relational agents in clinical psychiatry. Harv Rev Psychiatry 2010;18(2):119-130. [doi: 10.3109/10673221003707538] [Medline: 20235777]
76. Gaffney H, Mansell W, Tai S. Conversational agents in the treatment of mental health problems: mixed-method systematic review. JMIR Ment Health 2019 Oct 18;6(10):e14166 [FREE Full text] [doi: 10.2196/14166] [Medline: 31628789]
77. Bai YK, Lee S, Overgaard K. Critical review of theory use in breastfeeding interventions. J Hum Lact 2019 Aug;35(3):478-500. [doi: 10.1177/0890334419850822] [Medline: 31100027]
78. World Population Review. 2020. High Income Countries 2020 URL: http://worldpopulationreview.com/countries/high-income-countries/ [accessed 2020-02-28]
79. UNICEF. 2010. Infant and Young Child Feeding Programme Review: Consolidated Report of Six-country Review of Breastfeeding Programmes URL: https://www.unicef.org/nutrition/files/IYCF_Booklet_April_2010_Web.pdf [accessed 2020-05-13]

Abbreviations

BSES-SF: breastfeeding self-efficacy scale—short form
e-technologies: electronic technologies
ICT: information and communications technology
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
WHO: World Health Organization

Edited by G Eysenbach; submitted 09.12.19; peer-reviewed by R Khajouei, C Reis, R Bensley; comments to author 13.01.20; revised version received 05.04.20; accepted 27.04.20; published 29.05.20.

Please cite as:
Almohanna AA, Win KT, Meedya S
Effectiveness of Internet-Based Electronic Technology Interventions on Breastfeeding Outcomes: Systematic Review
J Med Internet Res 2020;22(5):e17361
URL: http://www.jmir.org/2020/5/e17361/
doi:10.2196/17361
PMID:32469315

Original Paper

Toward a Risk-Utility Data Governance Framework for Research Using Genomic and Phenotypic Data in Safe Havens: Multifaceted Review

Kerina Jones[1], BSc, PhD; Helen Daniels[1], BSc, PhD; Sharon Heys[1], BSc; Arron Lacey[1], BSc, PhD; David V Ford[1], MBA

Population Data Science, Swansea University Medical School, Swansea University, Swansea, United Kingdom

Corresponding Author:
Kerina Jones, BSc, PhD
Population Data Science
Swansea University Medical School
Swansea University
Population Data Science, Swansea University Medical School,
Singleton Park
Swansea, SA28PP
United Kingdom
Phone: 44 01792602764
Email: k.h.jones@swansea.ac.uk

Abstract

Background: Research using genomic data opens up new insights into health and disease. Being able to use the data in association with health and administrative record data held in safe havens can multiply the benefits. However, there is much discussion about the use of genomic data with perceptions of particular challenges in doing so safely and effectively.

Objective: This study aimed to work toward a risk-utility data governance framework for research using genomic and phenotypic data in an anonymized form for research in safe havens.

Methods: We carried out a multifaceted review drawing upon data governance arrangements in published research, case studies of organizations working with genomic and phenotypic data, public views and expectations, and example studies using genomic and phenotypic data in combination. The findings were contextualized against a backdrop of legislative and regulatory requirements and used to create recommendations.

Results: We proposed recommendations toward a risk-utility model with a flexible suite of controls to safeguard privacy and retain data utility for research. These were presented as overarching principles aligned to the core elements in the data sharing framework produced by the Global Alliance for Genomics and Health and as practical control measures distilled from published literature and case studies of operational safe havens to be applied as required at a project-specific level.

Conclusions: The recommendations presented can be used to contribute toward a proportionate data governance framework to promote the safe, socially acceptable use of genomic and phenotypic data in safe havens. They do not purport to eradicate risk but propose case-by-case assessment with transparency and accountability. If the risks are adequately understood and mitigated, there should be no reason that linked genomic and phenotypic data should not be used in an anonymized form for research in safe havens.

(J Med Internet Res 2020;22(5):e16346) doi:10.2196/16346

KEYWORDS

genomic data; data safe havens; data governance

Introduction

Background

The use of genomic data to revolutionize health research and clinical care is a major expanding area of investigation and development. Research using genomic data opens up new insights into health and disease to inform population health [1] and to develop precision medicine, namely, treatment based on a person's biological constitution, lifestyle, and environment [2]. Being able to use genomic data in association with health and administrative record data can multiply the benefits by including information such as comorbidities, medication histories, laboratory tests, education records, social issues, and lifestyle factors. This creates a more rounded and realistic picture to avoid genetic determinism and to ground phenotypically observed phenomena. Genomic data can be defined as the totality of a person's DNA sequence, and genetic data can be defined as the parts of the DNA that code for genes [3]. In this paper, we will refer to genomic data unless we are specifically mentioning genetics. We will refer to health and administrative data as phenotypic data for convenience, while acknowledging that not all the information in these records is ultimately or attributably genomic.

By their nature, genomic data are commonly considered to be among the more sensitive types of personal data for various reasons. These include their persistence through a person's life (barring a degree of plasticity), their role in predicting disease onset or likelihood, stigmatization for insurance or employment, and impact on kin [4,5]. This has led some to the concept of *genetic exceptionalism*, proposing that genetic data should be subject to new, particularly stringent use restrictions, whereas others argue that, although careful control measures are needed, the fundamental issues are largely the same as for other health data [6]. At the other end of the perception scale, some believe that all human genome sequence arising from publicly funded research should be freely available in the public domain [2]. Crucially, the ability of genomic data to provide in-depth information about a person is the defining factor in their added value for research and clinical care. Debates continue, and we have a conundrum in terms of how to progress the use of genomic data for public benefit while safeguarding individuals from harm.

We believe that there is a need for guidance toward a risk-utility data governance framework to simultaneously mitigate disclosure risk and retain maximum data utility. This is because of the nature of the data and ongoing discussions about the ability to truly anonymize genomic data. Some of this debate arises from a conflation of unique versus identifiable data, which we will return to later in this paper. In the absence of absolute certainty about the anonymization of genomic data beyond a shadow of a doubt, we propose that the framework needs to be practical, enabling research while safeguarding against all manageable risks. In this endeavor, privacy protectionism must be avoided, that is, the application of superfluous control measures that do not enhance privacy but do damage data usefulness [7]. Commonly used nonperturbative disclosure controls include suppression of variables or entire records, aggregation (such as age into bands), and masking (such as of clinic identifiers to protect professional reputation). Further methods include data perturbation (such as variable swapping between records), homomorphic encryption (a technique enabling computations on encrypted data), and other privacy-enhancing technologies [8]. Many of these are still in developmental stages and might not ultimately prove suitable for real-world application without reducing data utility. Even when algorithms are brought to encrypted data, they might not be immune to reidentification risk [9,10].

In a previous study, we explored the views of the general public on access models for genomic data in conjunction with health data [11]. Because so far genomic analyses have not yet revolutionized medical care and the majority of testing in clinical routine is still conventional, the study focused on genomic data that had been primarily collected for research. The 3 models of access presented were open (such as via a website), released externally to approved researchers, and within a data safe haven. We refer to a data safe haven as an infrastructure enabling data access within a secure environment subject to procedural, technical, and physical controls, in combination with disclosure controls applied to the data, to safeguard data subject privacy [12]. Although some people were comfortable with data being open access and more with release to named researchers, the overall preference was for access via data safe havens [11]. It can be argued that dangers differ between researcher-hosted data being unintentionally released externally leading to disclosure, compared with data hosted in a safe haven being reengineered to reveal personal data, as the latter is generally perceived to be more computationally intensive. Even so, because genomic data (largely) persist over the course of a lifetime and are passed to the offspring, it is important to remember that what is impossible today may be achievable within minutes in the not-too-distant future. Nonetheless, at this point in time, we acknowledge the appropriateness of all 3 models in their place and ongoing work to strengthen governance arrangements, for example, in Web-based access to large-scale genomic sequence for genome-wide associations and similar studies [13].

Objective

Many countries worldwide are investing in infrastructures that enable extensive, population phenotypic data to be accessed in an anonymized form via a data safe haven, and some are also incorporating genomic data (in various forms) to add to the research potential [14-16]. Being able to link data at the individual level while safeguarding privacy is an essential part of these enterprises, enabling information from multiple disparate datasets to be used in research. However, because of the nature of, and perceptions around, genomic data, there are challenges to be overcome to incorporate genomic data in a lawful, ethical, and socially acceptable way. Consequently, the aim of this paper was to work toward a risk-utility data governance framework for using genomic data in conjunction with phenotypic data in an anonymized form for research in data safe havens. We propose that the findings of this study will have particular value in developing guidelines for safe havens and augmenting existing operating models and will also be relevant to some extent for other data access models.

Methods

We took a multifaceted approach drawing upon data governance arrangements in published research papers, case studies of organizations working with genomic and phenotypic data, public views and expectations for the use of genomic data, and example studies using genomic and phenotypic data. The findings were contextualized against a backdrop of legislative and regulatory requirements and used to create a set of recommendations to inform a risk-utility data governance framework for using genomic and phenotypic data in safe havens. We set out our considerations of these elements in the Results section to draw together the findings to inform the recommendations. Ethical approval was not required for this study because the engagement with members of the public as research participants took place in our previous study [11].

Legislation and Literature Review

We carried out a summary (nonexhaustive) review of the legislative and regulatory backdrop to gain an insight on issues associated with the use of genomic data in principle. For this, we focused on the European Union (EU) because wider consideration was beyond the scope of this paper and would warrant a separate study. The elements considered were the European Convention on Human Rights and Biomedicine [17] and the General Data Protection Regulation (GDPR) [18] of 2016, augmented by national legislation and official guidance, such as the UK Data Protection Act of 2018 [19] and the work of the UK Information Commissioner's Office. We appreciate that legislation in other jurisdictions will vary, but we use the ones we have mentioned here as illustrations. In addition to compliance with data protection legislation, there are ethical issues to consider and address. The Global Alliance for Genomics and Health (GA4GH) has several working groups focusing on particular issues and has produced a high-level framework for the use of genomic and health data [20] to be implemented in line with jurisdictional requirements.

We used our previously conducted literature review on uses of genomic and routinely collected phenotypic data to guide this study (Daniels H et al, unpublished data, 2019). As part of the review, we drew out various pieces of information to categorize the studies and the details of data governance arrangements to the extent that they were presented in the publication. The relevant variables were genomic data and source, phenotypic data and source, data access model, and details of data governance. This information enabled us to create examples of data use to consider relative risks for data access via a safe haven. We did this using factors including the format of genomic data being used, the type of health condition being studied, and the extent of data linkage. We note that many of the published studies used an external data release model, but we used them in this context simply as examples of research using genomic and phenotypic data. We also drew upon a series of interviews we had conducted with representatives from safe haven enterprises in Germany, Australia, the United Kingdom, and Canada to add depth to the information in published studies. The relevant questions were on types of genomic data integrated with phenotypic data, main governance challenges encountered and how they were addressed, main access model, and access conditions [21].

Public Views

Recognizing that public engagement on the use of health data and specifically genomic data is an active area of investigation [22], when we carried out public engagement on the use of genomic and health data, we focused specifically on access models, as noted above [11]. The public engagement activities took the form of 8 workshops with a total of 116 people. As well as taking part in free-form discussions, the participants completed an anonymous exit questionnaire to provide their personal views on the relative benefits and risks of each access model: open, external release, or data safe haven. This had previously not been explored and provided new insights into public preferences to inform this paper.

Results

Legislative and Regulatory Backdrop

The EU has expressly recognized the use of genetic information in the European Convention on Human Rights and Biomedicine (1997), which provides for the misuse of biological and medical advances and sets out prohibitions in respect of bioethics and the right to a private life [17]. It further bans decisions made on the basis of genetic characteristics and governs predictive genetic tests for medicinal purposes. Personal data processing is largely governed by the GDPR [18] of 2016, augmented by national legislation and official guidance, such as the UK Data Protection Act of 2018 [19] and the work of the Information Commissioner's Office. Personal data are defined under Article 4 of the GDPR as "any information relating to an identified or identifiable natural person who can be identified, directly or indirectly in particular by reference to an identifier such as name...or to one or more factors specific to the physical, physiological, genetic specific to that natural person." Recital 34 of the GDPR defines genetic data as "personal data relating to the inherited or acquired genetic characteristics of a natural person which result from the analysis of a biological sample from the natural person in question, in particular chromosomal, deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) analysis or from the analysis of another element enabling equivalent information to be obtained." Processing personal data related to the health of an individual is provided for by the explicit provisions of GDPR Article 9. This provides differing lawful bases for processing and covers a range of data that are considered to be sensitive. Genetic data are expressly included within Article 9.

The key point in applying data protection legislation is that it applies to identifiable personal data from living individuals and not to data from which a person cannot be identified. Although the majority of the genome is identical in all humans, sequence variants occur across the genome that can be used to characterize people and groups. That is not to say that they can be used to identify an individual, as, although the information may be unique, this does not necessarily render it legally identifiable. Within the remit of the GDPR, data from which an individual cannot be identified are legally anonymous. However, identifiers as defined by law do not have to be direct and can be formed

from a number of pieces of information. As with any data source, genetic data linked to other data may give rise to a greater risk of identity disclosure, such that it is important to consider whether the combination of available data could lead to the identification of an individual. For example, Y-chromosome repeats were mined from the 1000 Genomes database and could identify people by cross-referencing them with ancestry databases [20]. As far as the GDPR is concerned, it is not correct to state that genetic data are always personal data or will always carry a high risk of disclosure. A proper legal approach would be to assess each use case with expert input to adequately understand the risk of identifiability. Applying legal governance to genetic data can be perceived as challenging because of the esoteric nature of the data, which, furthermore, can be shrouded in scientific terminology. This can appear to complicate the legal issues; however, in essence, data protection law governing genetic data is largely the same as for other health data types.

The GA4GH framework for the use of genomic and health data [21] is to be implemented in line with jurisdictional requirements. The framework establishes a set of foundational principles, namely, to respect individuals, families, and communities; advance research and scientific knowledge; promote health, well-being, and the fair distribution of benefits; and foster trust, integrity, and reciprocity. Furthermore, it proposes core elements of responsible data sharing: transparency; accountability; engagement; data quality and security; privacy, data protection, and confidentiality; risk-benefit analysis; recognition and attribution; sustainability; education and training; and accessibility and dissemination. It does not relate specifically to data safe havens and is broader in scope than our aim to work toward a risk-utility data governance framework for research using genomic data in conjunction with phenotypic data for research in safe havens. However, this valuable document formed a part of the backdrop and, along with the legislation, informed the development of the recommendations.

Forms of Genomic Data

The literature review on the use of genomic data with phenotypic data revealed a plethora of studies on many health conditions and using various forms of genomic data (Daniels H et al, unpublished data, 2019). Genome-wide association studies (GWAS) and phenome-wide association studies (PheWAS) use statistical methods applied to genomic sequence data to explore and correlate variants and phenotypic traits [22]. However, further processing often takes place so that data derivatives, or metadata, can be taken forward for use with phenotypic data for research or to inform clinical care. This is important when considering data governance issues because not all genomic data used in research are composed of sequence. In fact, the resulting metadata can take many forms of varying complexity. Some examples are as follows: Binary Alignment Map files, a compression of the sequence that can be annotated to explain particular details; Variant Call Format (VCF) files, providing information on the type, number, and position of nucleotide variants; single-nucleotide polymorphism (SNP) files detailing changes in single base pairs in the DNA sequence; risk score files, which contain information on risks of health conditions based on single or multiple genes; and among the simplest are files that detail the presence or absence of a trait of interest (Daniels H et al, unpublished data, 2019). Using genomic information in conjunction with additional data can add valuable detail and provide context in health conditions. However, importantly, the relative risks will differ depending on the form of genomic data being used, along with other factors. We take this into account in the development of the recommendations.

Examples Using Genomic and Phenotypic Data

We provide examples of studies using various forms of genomic data and phenotypic data to illustrate data use for research with a view to considering relative risks. As well as GWAS and PheWAS studies that require the full genome sequence for large-scale statistical analysis to identify variants of interest, many studies begin with sequence but subject it to further processing to create derivatives for linkage to structured phenotypic data. For example, polygenic risk scores derived from GWAS and SNPs then linked to an index of deprivation and individual postcodes to explore factors influencing alcohol dependence [23]. Genotyping at particular loci has been used with the number and duration of in-patient events in schizophrenia [24]. Gene expression profiling has been used in conjunction with electronic health records (EHRs) to compare breast cancer treatments and predict chemotherapy efficacy and outcomes across health care systems [25].

SNPs are among the forms of genomic data commonly used in conjunction with phenotypic data. These were used in a study on major depressive disease with linkage to health service data, including diagnoses, history of antidepressant prescribing, and referrals to secondary care for specialist treatment [26]. A study on dementia used a set of 6 SNPs linked to EHRs to establish and monitor the dementia status of participants [27]. A further example is the use of SNPs to study herpes zoster linked to EHRs to determine diagnosis in adults [28]. EHRs are often the source of phenotypic data, but information may also be drawn from registries [29,30], and in some cases, area-based measures, such as indices of deprivation, are included [27].

As for any studies using individual-level health data, risks will vary depending on the specifics of the use case. In relation to genomic data with phenotypic data, variables that may influence risk include forms of genomic data, common versus rare conditions, studying sensitive or stigmatizing conditions, and the extent of data linkage to phenotypic data. These are in addition to demographic factors such as extremes of age and small geographies, and all may combine to create the risk profile. As such, the degree of disclosure risk varies, but the repercussions of reidentification also vary. For example, the unauthorized disclosure that someone's close blood relative carries a high risk of developing a strongly hereditary condition such as Huntington disease will have serious implications for themselves and their kin. On the other hand, conditions such as heart disease and diabetes are multifactorial and, as such, are more complex and indefinite in terms of disease prediction. These and a myriad of other considerations lead us to propose that the risks and benefits of planned data uses should be taken into account on a case-by-case basis.

Data Governance Arrangements

We used information from the literature review and the series of interviews with representatives of safe havens on data governance arrangements in place. We refer only to governance arrangements in relation to the use of genomic data, rather than the work of these enterprises in general. We found that differing solutions have been put in place to enable genomic data to be used in conjunction with phenotypic data. Examples are given here to show variety.

The German Medical Informatics Initiative is a major infrastructural investment that has created multiple university hospital consortia to integrate clinical data, including genomics [31]. In terms of genomic data, the work of this new initiative is beginning with biosample collection before the results of genomic analyses will be shared across consortia. Germany is subject to the GDPR but also has stringent national privacy regimens for general and genomic data processing [32,33]. Moreover, apart from repositories holding data locally, methods for cross-center distributed analyses, such as DataSHIELD, are being employed [34]. These will bring the analysis to the data and avoid the need to share data where this is deemed unacceptable to oversight committees. Subject to approval, data can be used within consortia, across consortia, and by any approved researcher. However, as noted, this might not involve direct access to data but might depend on privacy-preserving analysis methods. There is a network of committees and working groups for the establishment of agreed standards.

The Sax Institute in Australia manages the *45 and up* study, a longitudinal cohort of over 250,000 people, including phenotypic and genomic data [35]. Sax works in partnership with the Garvan Institute of Medical Research, which acts as a genome sequencing facility and makes data available for research subject to conditions [36]. Sax operates as a repository for health and self-report data, whereas Garvan acts as a repository for the genomic data. Although the full genome data are generally too large to move, other data such as VCF files could be transported; thus, the decision to use a distributed model is based on data governance reasons. Researchers need to apply to both institutional data access committees before access to genomic and phenotypic data can be granted. Linked data can be accessed by public and private sector researchers via a portal subject to ethical and other relevant approvals. Following analysis, outcomes are released externally but not row-level data.

The UK Secure Research Platform (SeRP) is a data infrastructure housing the Secure Anonymised Information Linkage (SAIL) databank and various other initiatives [15,37]. UK SeRP can be customized to implement the data governance model required by particular programs of work. To date, SAIL has integrated genomic data on a project basis but is working to incorporate more genomic data as part of standard data feeds, along with phenotypic data from health and administrative records. Example projects involve a psychosis cohort and an epilepsy study. As the genomic data were brought into SAIL for particular projects, separate ethical approval including participant informed consent was obtained. The psychosis cohort brought in polygenic risk scores and copy number variants; the epilepsy study brought in VCF files. In both cases, the data were only made available to project researchers because of regulatory approvals. But in general, SAIL allows data access to any approved public sector researcher; the commercial sector must work with a SAIL analyst or another public organization to access data on their behalf. All proposals to access data for research must be approved by an information governance review panel that co-opts additional experts in assessing particular data types, such as genomic. Some data providers also reserve the right to review data use proposals, in addition to the panel.

The Institute for Clinical Evaluative Sciences (IC/ES) operates in Ontario, Canada [38], and holds a highly phenotyped cohort of over 2000 children and young people with a neurological development disorder, most frequently, autism. DNA samples were collected for whole genome sequencing from the entire cohort, and the ones who consented to linkage have their VCF file data linked to IC/ES health and administrative data. As such, this is a project-level development, but it is anticipated that linkage of genomic data to phenotypic data will become more routine in future. The privacy approval group at IC/ES was concerned about identifiability because the genomic data are often unique, particularly where there are rare variants. Owing to this, the arrangement at IC/ES has been to hold the genomic data separately and not on the main analysis platform. As the linked data are considered highly sensitive, only an IC/ES analyst has access. The project lead prepares queries to be executed on the data and receives the results from the IC/ES analyst.

From these examples, it can be seen that data safe havens are at different stages of development in terms of integrating genomic data, they are working with various formats of genomic data, and they have varying perceptions of risk leading to differing requirements for data access. These are just some examples, with others underway such as the Swiss Personalized Health Network and its associated BioMedIT project. Through this, there is a network of core facilities for the secure processing of biomedical data across Swiss universities, enabling research within a secure environment [39]. However, although models differ, one of the common beauties of safe havens (over open access or external release of data) is that they are able to apply a suite of disclosure controls directly to the data (in totality or at a project level) and across the whole environment, thus managing risk across all stages from data incorporation to archiving [14,15,38]. We used the information gained from these various models to guide the development of the recommendations.

Public Views

Having previously reported public perspectives on 3 models of accessing genomic data [11], here we draw upon the views gained on data use in safe havens, in line with the focus of this paper. In general, workshop participants were less concerned about the use of genomic data in safe havens than external release or open access. However, there were some provisos in relation to safe havens. Participants wanted to be properly informed on the purpose of data use and for analyses to be conducted by approved researchers, with concern expressed on misuse by commercial companies. They wanted to be reassured that appropriate safeguards would be in place, with data use

being auditable and controlled. The need for consent to reuse genomic data that had been primarily collected for research emerged strongly (for all access models) and led us to propose wording for the information sheet and consent form in prospective studies [11]. The use of anonymized data in safe havens was also seen as a way to mitigate risks of discrimination because access is limitable by systematic controls. The public viewpoints on the use of genomic data in safe havens guided us in the creation of recommendations.

Discussion

Principal Findings

Having considered the legislative backdrop in the EU, it can be seen that there is nothing inherently different in the EU GDPR about the lawful provisions for genomic data processing compared with other health data: all are classed as special category data. However, there is much debate about the effectiveness of anonymization processes for genomic data, and some consider that it is particularly difficult to produce genomic data that are both anonymous and useful [3,10]. This is not a new thought as it has long been noted in relation to demographic and health data [40], but it has been specifically applied because of the nature of genomic data leading to the concept of genetic exceptionalism [6]. Some of this is because of factors such as persistence, discrimination, and impact on kin, but there is an apparent mystique around genomic data that is proving challenging even if it is nonempirical, as beliefs play a significant role in policy making. The perceived complexities surrounding genomic material and data can lead to an overly cautious and proscriptive approach to the detriment of research [3]. We fully acknowledge that there are risks to be mitigated for safe data use and that data misuse can cause harm to individuals and professional reputations. However, it is also true that serious harm occurs because of data nonuse, far more than the missed benefits of proper data usage and leading to hundreds of thousands of lost lives and billions of dollars in financial burdens to societies [41]. In fact, it has been postulated that nonuse is often a greater problem than unauthorized data use [42].

A particularly important issue is the problem of conflating unique and identifiable. It cannot be overstressed that just because data are unique, this is not the same as being identifiable *per se*. If research had to rely on using only the data present in equivalence classes (ie, where there are no unique records, only sets of records [minimally 2] with identical variables), progress would be hampered. Even the fact that a person's full genome is unique does not render it identifiable without considerable effort, interpretation, and additional information to confirm identity. This has been shown in the debunking of a study purporting to identify individuals from genomic data, revealing that the work only narrowed down the information to a category of people based on phenotypic traits [43,44]. That is not to say that new insights on current limited knowledge of the genome will not change the likelihood of identifiability. Future proofing is important in the use of all personal data but more so in relation to genomics where successive new understandings are being revealed from previously unknown data content as the genomic revolution advances apace.

Another important factor to consider is what is meant by genomic data when special processing rules are suggested. As we have seen from the literature review and case study interviews, genomic data are used in various forms along with a variety of phenotypic data. Although GWAS and PheWAS studies rely on DNA sequence, many other studies use only DNA derivatives, or metadata, of varying complexity. We propose that it is not appropriate to bundle genomic data into one category and assume a uniformly high level of privacy risk. Even so, there is more than legislation to consider when seeking to use personal data for research, including ethical and social implications. Among the safe havens currently working with genomic data, there is a variety of approaches to data governance in accordance with their jurisdictional frameworks and interpretations. Depending on how the genomic data were collected, there may be a requirement for informed consent or data may be incorporated via standardized feeds along with health and administrative records based on sharing agreements with data providers. Data access might be limited to institutional employees or the public sector or might be open to all sectors. Researchers might be provided with direct access to linked data or to be limited to distributed queries. Genomic data are sometimes stored with the phenotypic data, but they might be on separate platforms and are sometimes held by separate organizations with access provided on a distributed model or with only limited derivative data imported. Commonly, safe havens do not generally release row-level data but export products of analysis following disclosure risk scrutiny. In terms of social acceptability, at least from our previous work, the public were generally favorable toward the use of genomic data in safe havens, subject to provisos including privacy safeguards and being provided with information about data uses.

Recommendations

We use the findings of the various elements of this study to propose a set of recommendations toward creating a risk-utility data governance framework and to augment existing operating models for using genomic and phenotypic data for research in an anonymized form in safe havens. We present the recommendations as overarching principles and practical control measures to mitigate risks and retain maximum data utility.

Principles

We propose that these principles be applied in general as part of the data governance framework for safe havens incorporating genomic data. We align the principles with the core elements presented in the high-level framework for the use of genomic and health data developed by the GA4GH [20], with the main relevant elements shown in italic text. The principles are as follows:

- Jurisdictional data protection legislation for general data processing, any specific provisions for genomic data processing, and relevant authoritative guidance and codes of practice should be examined and properly interpreted, with the input of legal expertise to ensure due compliance, *transparency,* and *accountability.*

- The need for regulatory approvals (including research ethics and informed consent) and data provider permissions for incorporating genomic data should be assessed, and all due diligence should be followed to demonstrate *transparency* and *accountability* and to support *recognition* and *attribution* in data provision.
- Members of the public should be provided with the opportunity to be properly *engaged* and involved at the strategic level to contribute their views to informing developments for the systematic incorporation of genomic data. The respectful need for a measure of *education and training* and the *dissemination* of information in *accessible* forms should be accommodated for transparency and accountability to the public, from whom the data arise.
- There is a need for a flexible, proportionate approach to consider proposed uses of genomic data, as they come in many formats with varying privacy risks, and (at least at present) it is incorrect to assume that genomic data are identifiable simply by reason of being unique. Assessment should be made on a case-by-case basis using *risk-benefit analysis* and with *education and training* to support data access committees.
- Safe havens should use their combination of data environment, physical, technical, and procedural controls, coupled with disclosure controls applied directly to the data, to maximize *data quality and security, privacy,* and *data protection and confidentiality.*
- As knowledge in the field of genomics is advancing rapidly, it is especially important to ensure future proofing of individual *privacy, data protection, and confidentiality* of genomic information. For *accountability* and *sustainability* of data use, we recommend that the practice of only exporting products of analysis should be the general policy, particularly for linked genomic and phenotypic data. Exceptions could be made if all relevant approvals are in place and the data are to be moved to another safe haven.

Practical Control Measures

The following is a selection of practical control measures that can be applied to specific cases where genomic and phenotypic data are used in combination to mitigate risk and retain maximum utility. They are not sequential or hierarchical but can be used in various combinations. The proposed control measures are as follows:

- All relevant project-level regulatory approvals for the collection and processing of genomic data should be checked by the data access committee, as part of their assessment of data use suitability.
- Researchers should consider opportunities for public engagement on their proposed use of genomic data to gain input on direction and preliminary findings.
- Along with the standing membership of the data access committee, a genomics expert should be co-opted to review particular proposals.
- Case-by-case review should take into account factors including the form and extent of genomic data to be used and the rarity and sensitivity of health and lifestyle factors to be studied, in addition to the criteria generally used by the data access committee.
- Reviews should include who will be permitted to access the data, depending on perceived risk and regulatory and data provider requirements. If deemed necessary, data access should be limited to the project team, a single researcher, or only to an analyst employed by the safe haven host, tasked with producing results to share with the project team.
- Many safe havens provide remote data access for approved researchers at their desktop, wherever they are based. If deemed necessary, data access should be physically restricted to a designated safe room so that a researcher has to be present on site when analyzing the data, and data access should be subject to stringent monitoring.
- Data granularity can be curtailed using a variety of disclosure control techniques applied to the data. The choice of methods should be selected with care in discussion with the researcher to retain maximum utility and safeguard the data.
- If available to the safe haven, the option of using distributed queries where the researcher is not provided with a view of the data but sends their query to the server and receives results should be considered for proposals deemed particularly risky.
- Where genomic data cannot be moved from the source, a hybrid model should be considered. Instead of incorporating the genomic data into the data safe haven, a federated access model may be used. However, the feasibility of this will depend on available technology and resources.

From our findings we anticipate that, in many cases, genomic data can be used safely and appropriately in conjunction with phenotypic data within a safe haven without many (if any) major changes to the current operating models, making use of the suite of controls available. The use of various forms of genomic data with phenotypic data will often be unlikely to present significant additional risks over and above those posed by the use of multiple linked health and administrative records within the safe haven. Provided that a case-by-case assessment is made, and proportionate controls are applied to mitigate risks while transparently acknowledging that they might not be totally eradicated, there should be no reason for not permitting the use of genomic data for research in safe havens.

What This Study Adds

This is the first known study to propose recommendations toward a risk-utility data governance framework for the use of genomic and phenotypic data in safe havens. It has brought together findings from published research, case studies of data safe havens, and public views against a backdrop of (EU) data protection legislation to inform the perspectives presented. It is a novel, evidence-based study that can be used to guide existing and newly developing data safe havens on working with genomic data to safeguard the data without falling into the trap of privacy protectionism [7], but still ensuring risks are properly mitigated while retaining maximum data utility. We do not claim to have solved all the challenges or that risk can be totally eradicated, but the study has shed new light on routes toward a risk-benefit data governance framework to use genomic and phenotypic data safely and effectively.

Limitations

We acknowledge limitations to this study. It is based on a nonexhaustive literature review, views of a limited number of people based on a variety of settings in Wales, case studies of some data safe havens, and the main EU data protection legislation. It is possible that various other pieces of information pertaining to other jurisdictions, organizations, and cultures may differ. However, we are not proposing that our findings are ultimately definitive but that they can be used toward a data governance framework, taking necessary differences into account.

Conclusions

This study acknowledges the benefits and challenges in using genomic data in conjunction with phenotypic data and the need for guidance to promote the safe, socially acceptable use of data in data safe havens. We used a multifaceted approach to propose evidence-based recommendations toward a risk-utility data governance framework based on a suite of controls applied to and around the data to mitigate risks and retain data utility. They do not purport to eradicate risk but propose case-by-case assessment with transparency and accountability. If the risks are adequately understood and mitigated, there should be no reason that linked genomic and phenotypic data should not be used in an anonymized form for research in safe havens.

Acknowledgments

This study received funding from the UK Medical Research Council (MC_PC_16035).

Conflicts of Interest

None declared.

References

1. Molster CM, Bowman FL, Bilkey GA, Cho AS, Burns BL, Nowak KJ, et al. The evolution of public health genomics: exploring its past, present, and future. Front Public Health 2018;6:247 [FREE Full text] [doi: 10.3389/fpubh.2018.00247] [Medline: 30234091]
2. Dickenson D, van Beers B, Sterckx S. Introduction in personalised medicine. In: Dickenson D, van Beers B, Sterckx S, editors. Personalised Medicine, Individual Choice and the Common Good. Cambridge: Cambridge University Press; 2018.
3. Finnegan T, Hall A. Identification and Genomic Data. Cambridge, UK: PHG Foundation; Dec 2017.
4. Green RC, Lautenbach D, McGuire AL. GINA, genetic discrimination, and genomic medicine. N Engl J Med 2015 Jan 29;372(5):397-399. [doi: 10.1056/NEJMp1404776] [Medline: 25629736]
5. Joly Y, Feze IN, Song L, Knoppers BM. Comparative approaches to genetic discrimination: chasing shadows? Trends Genet 2017 May;33(5):299-302. [doi: 10.1016/j.tig.2017.02.002] [Medline: 28365141]
6. Chin JJ, Campbell AV. What — if anything — is special about 'genetic privacy'? In: Sheung-hung KT, Ho CW, editors. Genetic Privacy: An Evaluation of the Ethical and Legal Landscape. Covent Garden, London , UK: Imperial College Press; 2013:223-251.
7. Allen J, Holman CD, Meslin EM, Stanley F. Privacy protectionism and health information: is there any redress for harms to health? J Law Med 2013 Dec;21(2):473-485. [Medline: 24597394]
8. Elliot M, Mackey E, O'Hara K, Tudor C. UK Anonymisation Network. 2016. The Anonymisation Decision-Making Framework URL: https://ukanon.net/wp-content/uploads/2015/05/The-Anonymisation-Decision-making-Framework.pdf [accessed 2018-11-30]
9. The Royal Society. Protecting Privacy in Practice. The Current Use, Development and Limits of Privacy Enhancing Technologies in Data Analysis. London, UK: The Royal Society; 2019.
10. Naveed M, Ayday E, Clayton EW, Fellay J, Gunter CA, Hubaux JP, et al. Privacy in the genomic era. ACM Comput Surv 2015 Sep;48(1):6 [FREE Full text] [doi: 10.1145/2767007] [Medline: 26640318]
11. Jones KH, Daniels H, Squires E, Ford DV. Public views on models for accessing genomic and health data for research: mixed methods study. J Med Internet Res 2019 Aug 21;21(8):e14384 [FREE Full text] [doi: 10.2196/14384] [Medline: 31436163]
12. Lea N, Nicholls J, Dobbs C, Sethi N, Cunningham J, Ainsworth J, et al. Data safe havens and trust: toward a common understanding of trusted research platforms for governing secure and ethical health research. JMIR Med Inform 2016 Jun 21;4(2):e22 [FREE Full text] [doi: 10.2196/medinform.5571] [Medline: 27329087]
13. Dyke S, Linden M, Lappalainen I, de Argila JR, Carey K, Lloyd D, et al. Registered access: authorizing data access. Eur J Hum Genet 2018 Dec;26(12):1721-1731 [FREE Full text] [doi: 10.1038/s41431-018-0219-y] [Medline: 30069064]
14. Ark T, Kesselring S, Hills B, McGrail K. Population Data British Columbia: a data resrouce for research. Int J Popul Data Sci 2019:- (forthcoming)(forthcoming).
15. Jones K, Ford D, Thompson S, Lyons R. The SAIL Databank on the UK Secure Research Platform. Int J Popul Data Sci 2020:- (forthcoming).

16. Flack F, Smith M. The population health research network - population data centre profile. Int J Popul Data Sci 2019 Nov 20;4(2):- (forthcoming). [doi: 10.23889/ijpds.v4i2.1130]
17. Council of Europe. Conseil de l'Europe. 1997. Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine URL: https://rm.coe.int/168007cf98 [accessed 2019-11-30]
18. General Data Protection Regulation (GDPR). 2016. URL: https://gdpr-info.eu/ [accessed 2019-09-30]
19. Government of UK Legislation. Data Protection Act 2018 URL: http://www.legislation.gov.uk/ukpga/2018/12/contents/enacted [accessed 2019-10-30]
20. The Global Alliance for Genomics and Health (GA4GH). 2014 Sep. Framework for Responsible Sharing of Genomic and Health-Related Data URL: https://www.ga4gh.org/wp-content/uploads/Framework-Version-10September2014.pdf [accessed 2019-05-30]
21. Global Alliance for Genomics and Health. GenomEthics. 2019. Your DNA, Your SAY URL: https://surveys.genomethics.org/survey/yourdnayoursay/en [accessed 2019-06-01]
22. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 years of GWAS discovery: biology, function, and translation. Am J Hum Genet 2017 Jul 6;101(1):5-22 [FREE Full text] [doi: 10.1016/j.ajhg.2017.06.005] [Medline: 28686856]
23. Clarke TK, Smith AH, Gelernter J, Kranzler HR, Farrer LA, Hall LS, et al. Polygenic risk for alcohol dependence associates with alcohol consumption, cognitive function and social deprivation in a population-based cohort. Addict Biol 2016 Mar;21(2):469-480 [FREE Full text] [doi: 10.1111/adb.12245] [Medline: 25865819]
24. Wickramasinghe A, Tulloch AD, Hayes RD, Chang CK, Broadbent M, Di Forti M, et al. Associations between the schizophrenia susceptibility gene ZNF804A and clinical outcomes in psychosis. Transl Psychiatry 2015 Dec 15;5:e698 [FREE Full text] [doi: 10.1038/tp.2015.198] [Medline: 26670283]
25. Afghahi A, Mathur M, Thompson C, Mitani A, Rigdon J, Desai M, et al. Use of gene expression profiling and chemotherapy in early-stage breast cancer: a study of linked electronic medical records, cancer registry data, and genomic data across two health care systems. J Oncol Pract 2016 Jun;12(6):e697-e709 [FREE Full text] [doi: 10.1200/JOP.2015.009803] [Medline: 27221993]
26. Hall LS, Adams MJ, Arnau-Soler A, Clarke TK, Howard DM, Zeng Y, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, et al. Genome-wide meta-analyses of stratified depression in Generation Scotland and UK Biobank. Transl Psychiatry 2018 Jan 10;8(1):9 [FREE Full text] [doi: 10.1038/s41398-017-0034-1] [Medline: 29317602]
27. Almeida OP, Morar B, Hankey GJ, Yeap BB, Golledge J, Jablensky A, et al. Longevity Klotho gene polymorphism and the risk of dementia in older men. Maturitas 2017 Jul;101:1-5. [doi: 10.1016/j.maturitas.2017.04.005] [Medline: 28539162]
28. Crosslin DR, Carrell DS, Burt A, Kim DS, Underwood JG, Hanna DS, et al. Genetic variation in the HLA region is associated with susceptibility to herpes zoster. Genes Immun 2015;16(1):1-7 [FREE Full text] [doi: 10.1038/gene.2014.51] [Medline: 25297839]
29. Shen L, Walter S, Melles RB, Glymour MM, Jorgenson E. Diabetes pathology and risk of primary open-angle glaucoma: evaluating causal mechanisms by using genetic information. Am J Epidemiol 2016 Jan 15;183(2):147-155 [FREE Full text] [doi: 10.1093/aje/kwv204] [Medline: 26608880]
30. Millwood IY, Bennett DA, Holmes MV, Boxall R, Guo Y, Bian Z, China Kadoorie Biobank Collaborative Group. Association of CETP gene variants with risk for vascular and nonvascular diseases among Chinese adults. JAMA Cardiol 2018 Jan 1;3(1):34-43 [FREE Full text] [doi: 10.1001/jamacardio.2017.4177] [Medline: 29141072]
31. Medizininformatik-Initiative: Vernetzen. Forschen. Heilen. 2019. URL: http://www.medizininformatik-initiative.de/en [accessed 2019-11-30]
32. Molnár-Gábor F. Germany: a fair balance between scientific freedom and data subjects' rights? Hum Genet 2018 Aug;137(8):619-626 [FREE Full text] [doi: 10.1007/s00439-018-1912-1] [Medline: 30116955]
33. European Commision. JRC Publications Repository. 2018. JRC Science for Policy Report: Overview of EU National Legislation on Genomics URL: http://publications.jrc.ec.europa.eu/repository/handle/JRC113479 [accessed 2019-09-30]
34. Wilson RC, Butters OW, Avraam D, Baker J, Tedds JA, Turner A, et al. DataSHIELD – new directions and dimensions. Data Sci J 2017;16:21 [FREE Full text] [doi: 10.5334/dsj-2017-021]
35. Sax Institute. 2019. 45 and Up Study URL: https://www.saxinstitute.org.au/our-work/45-up-study/ [accessed 2019-11-30]
36. Garvan Institute of Medical Research. 2019. URL: https://www.garvan.org.au/ [accessed 2019-11-30]
37. Jones KH, Ford DV, Thompson S, Lyons R. A Profile of the SAIL Databank on the UK Secure Research Platform. Int J Popul Data Sci 2019;4(2):4 [FREE Full text] [doi: 10.23889/ijpds.v4i2.1134]
38. Institute for Clinical Evaluative Sciences (ICES). 2019. URL: https://www.ices.on.ca/About-ICES [accessed 2019-11-30]
39. SIB Swiss Institute of Bioinformatics: Homepage. 2020. A National Infrastructure Network to Enable Secure Biomedical Data Processing: The BioMedIT Project URL: https://www.sib.swiss/about-sib/news/10337-a-national-infrastructure-network-to-enable-secure-biomedical-data-processing-the-biomedit-project [accessed 2019-11-30]
40. Ohm P. Electronic Privacy Information Center. 2009. Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization URL: http://epic.org/privacy/reidentification/ohm_article.pdf [accessed 2019-11-30]

41. St Clair D. Managed Healthcare Executive. 2008. Non-use of Patient Clinical Data a Greater Risk than Misuse URL: http://managedhealthcareexecutive.modernmedicine.com/managed-healthcare-executive/news/non-use-patient-clinical-data-greater-risk-misuse?page=full [accessed 2019-09-01]
42. Jones K, Laurie G, Stevens L, Dobbs C, Ford D, Lea N. The other side of the coin: Harm due to the non-use of health-related data. Int J Med Inform 2017 Jan;97:43-51 [FREE Full text] [doi: 10.1016/j.ijmedinf.2016.09.010] [Medline: 27919394]
43. Lippert C, Sabatini R, Maher MC, Kang EY, Lee S, Arikan O, et al. Identification of individuals by trait prediction using whole-genome sequencing data. Proc Natl Acad Sci USA 2017 Sep 19;114(38):10166-10171 [FREE Full text] [doi: 10.1073/pnas.1711125114] [Medline: 28874526]
44. Erlich Y. Major flaws in 'Identification of individuals by trait prediction using whole-genome sequencing data'. bioRxiv 2017:- preprint [FREE Full text] [doi: 10.1101/185330]

Abbreviations

EHR: electronic health record
EU: European Union
GA4GH: Global Alliance for Genomics and Health
GDPR: General Data Protection Regulation
GWAS: genome-wide association study
IC/ES: Institute for Clinical Evaluative Sciences
PheWAS: phenome-wide association study
SAIL: Secure Anonymised Information Linkage
SeRP: Secure Research Platform
SNP: single-nucleotide polymorphism
VCF: Variant Call Format

Edited by G Eysenbach; submitted 20.09.19; peer-reviewed by A Leichtle, J Brenas; comments to author 12.12.19; revised version received 13.01.20; accepted 30.01.20; published 15.05.20.

Please cite as:
Jones K, Daniels H, Heys S, Lacey A, Ford DV
Toward a Risk-Utility Data Governance Framework for Research Using Genomic and Phenotypic Data in Safe Havens: Multifaceted Review
J Med Internet Res 2020;22(5):e16346
URL: https://www.jmir.org/2020/5/e16346
doi:10.2196/16346
PMID:32412420

Viewpoint

All Our Data Will Be Health Data One Day: The Need for Universal Data Protection and Comprehensive Consent

Christophe Olivier Schneble[1], MSc; Bernice Simone Elger[1], Prof Dr; David Martin Shaw[1], PhD

Institute for Biomedical Ethics, University of Basel, Basel, Switzerland

Corresponding Author:
Christophe Olivier Schneble, MSc
Institute for Biomedical Ethics
University of Basel
Bernoullistrasse 28
Basel, 4056
Switzerland
Phone: 41 797252512
Email: christophe.schneble@unibas.ch

Abstract

Tremendous growth in the types of data that are collected and their interlinkage are enabling more predictions of individuals' behavior, health status, and diseases. Legislation in many countries treats health-related data as a special sensitive kind of data. Today's massive linkage of data, however, could transform "nonhealth" data into sensitive health data. In this paper, we argue that the notion of health data should be broadened and should also take into account past and future health data and indirect, inferred, and invisible health data. We also lay out the ethical and legal implications of our model.

(J Med Internet Res 2020;22(5):e16879) doi:10.2196/16879

KEYWORDS

big data; health data; social media; data protection; guidelines; best practices

Introduction

Background

Data intensive software, such as social media, wellness, and mobile health (mHealth) apps, have become ubiquitous in everyday life and are frequently used in a variety of situations. Years ago, social media networks were mostly accessed from traditional computers, but the rising use of smartphones and apps to access those networks has opened a Pandora's Box regarding data collection, including geolocation, motion data, health-related data, and behavioral data [1]. The collection of additional behavioral data about users was initially very limited, and only a fraction of basic data was collected (eg, IP address, operating system, and browser version). In contrast, current apps on smartphones have begun continuous monitoring of users by harvesting geolocation and motion data, and thus, they have the ability to infer users' physical and mental health states, for example, to detect signs of depression [2,3] and predict their next likely location [4,5]. Moreover, app companies have collected a tremendous amount of data on individuals' public and private activities in the digital world, which are being reused not only for the sake of their primary platforms, but also in other lucrative business sectors, such as robotics, life sciences, car manufacturing, and health data provision. Previously, users were able to simply opt out of these services, but this is becoming increasingly challenging nowadays given the monopoly market structure instilled by the companies that drive digital transformation. Indeed, customers are increasingly forced to use these services because they either do not have any other equivalent alternatives in terms of services provided or they are influenced by their peers or parents (in case of children) to use the services. Sometimes these companies nudge users with marketing strategies, such as substantial advantages and discounts offered only on these platforms. This proves to be problematic since to use these now important services, at least to some extent, users have to consent to some mandatory data sharing [6] and consequently expose their privacy.

Ambiguous Terms and Conditions

In view of the aforementioned facts, it is vital that the relationship between users and companies is transparent and regulated. This relationship is currently mostly defined in the terms and conditions (T&Cs), terms of service, and data privacy notices, which are unfortunately lacking in several aspects in term of enabling potential users to make an informed decision when signing up for a service. For instance, users are not warned about possible harms that might result from their activities on

the platforms (eg, linkage of several anonymous data sources could lead to reidentification of otherwise anonymous datasets and lack of awareness of secondary use could undermine user privacy and confidentiality). Additionally, the information provided in the T&Cs of different platforms is not reader-friendly and not succinctly summarized to nudge users to read them thoroughly. They are also not harmonized in the sense that each platform has its own implementation or they are simply not prominent enough during the process of signing up for the service [7,8]. Another weakness of T&Cs is that they do not make it clear that social media and the linkage of several independent unique databases can yield health data. Health data represent a special data category [9], requiring special security and privacy policies for governance. Indeed, many international legislations define health data as special data needing more protection than "usual" nonhealth-related data. The flipside of current legislation is that nonhealth-related data is subject to less strict governance. In the era where a digital phenotype [10] is emerging, data linkage can be very predictive and can be used to, for example, derive personal traits [11], predict psychosomatic diseases [2], and obtain other types of behavioral information.

Aim of the Paper

This paper presents a new approach for considering data, with the four categories of *direct*, *indirect*, *inferred*, and *invisible health* data, and suggests different types of possible consent frameworks that are up to this challenge, especially as data might not be conceived as distinct health data when produced. We first describe the already recognized categories of direct and indirect health data and then present the other two categories, describe the legal framework in the European Union and United States, and explore the different potential consent mechanisms and their suitability for these four categories of data.

Ubiquity of Health Data

Evolution of Health Data

Until the end of the last decade, health data were easily defined, and they included medical records, diagnostic images, laboratory testing data, and data produced by biomedical or clinical means. However, as rightly pointed out by Vayena et al [12], the notion of what is considered health data has considerably evolved. So-called biomedical big data nowadays ranges from data produced by health services, public health activities, and biomedical research to data registering exposure to environmental factors, such as sunlight and pollution, or data revealing lifestyle, socioeconomic conditions, and behavioral patterns, such as those from wellness and fitness apps, social media, and wearable devices. There is thus a paradigm shift from the notion of individual data producers and distinct categories to a more complex notion of a data ecosystem [13].

Implications of Massively Interlinked Data

The massive amount of data produced and interlinked has an effect on the characterization of individuals today, including their behavioral profile. Jain et al [10] developed the concept of the *digital phenotype*. The digital phenotype, an enlarged notion of the extended phenotype, encompasses the ubiquity of digital technologies and linkage of their data to virtually any other data, possibly resulting in potentially health-relevant data [14]. This notion is underpinned by not only the large numbers of studies conducted by universities that take advantage of the ubiquity of massive amounts of publicly available data, but also research involving the use of apps to predict individual behavior [15].

Further complicating these issues is the possibility that other types of data could be health data one day [16]. For example, some social media data concern exercise, which is highly relevant to health. Most people would agree that exercise data are health-related data and represent an example of indirect health data [10]. However, it is less obvious that other data, such as address and shopping data, location data, smart home data (eg, Amazon Alex and Siri data), smart car data, and articles shared on social media, can be combined with other datasets to infer health data [1,11], and this needs to be made more transparent in the future. It is already possible to use data to infer the degree of exposure to pollution and its likely health effects over long periods, and driving behavior data, such as acceleration patterns, can indicate the risks people take, which again could be used to infer their health conditions when combined with other elements.

Digitalization of Past Paper-Based Data

At another level, digitalization of past paper-based data could also yield direct or indirect health data (either through digitalization of paper medical records or fitness diaries), and future technologies, such as machine learning, may be able to identify health-relevant uses for data that have not yet even been conceived.

Thus, the days of health data as a distinct category are numbered, and soon, we will have not only direct health data, but also what we term indirect health data, inferred health data, and currently invisible health data (where the relevance to health might be perceptible only by machine intelligence in the future). This new understanding indicates the need for new governance, compliance, and regulatory mechanisms to handle data, protect individuals' privacy, and uphold the security of such new sensitive data. We briefly examine the current legislation that is relevant to these different categories of health data before continuing our ethical analysis.

Legal Situation in the United States and European Union

In contrast to the European Union, the United States is not subject to a single overarching data protection law. Data protection issues are implemented at the federal and state level. Table 1 summarizes some of the major federal laws that deal with data protection issues.

Table 1. Overview of the most important federal laws on data protection in the United States.

Law	Scope	Main points
Gramm Leach Billey Act	Governs protection of the personal information in the hands of banks, insurances companies, and other companies in the financial service industry.	Addresses "nonpublic information" (NPI) that institutions collect from their customers in connection with the provision of services. Imposes requirements for securing NPI, restricting disclosures, and using NPI. Obligation to notify customers when NPI is improperly exposed.
Fair Credit Reporting Act	Federal law regulating the collection of consumer credit information and access to credit card reports.	Governs how credit bureaus can collect and share information about individual consumers. Businesses check credit reports for many purposes, such as deciding whether to provide a loan or sell insurance to a consumer. This act also gives consumers certain rights, including free access to their own credit reports.
Health Information Portability Act	Protects information held by covered entities concerning health status, provision of health care, or payment for health care.	Breach notification. Data handling by covered entities and definition of safeguards.
Telephone Consumer Protection Act	The order is relevant to any company that uses automated technology to phone or send text messages to consumers.	Regulates telemarketing and forces companies to respect do not call registries.
Family Educational Rights and Privacy Act	Federal law that protects the privacy of student education records.	Offers students the right to correct information about themselves.

In the United States, some privacy frameworks provide a different definition of personal data and the notion of sensitive personal data varies among several federal state laws, with the Californian legislation being the most comprehensive, and among different economic sectors. Health data regulation is mostly included in the Health Insurance Portability and Accountability Act (HIPAA), which only covers entities that are directly related to health care operations, such as health care providers, health plans, and health care clearing houses, and the statement is as follows: "Except as otherwise provided, the standards, requirements, and implementation specifications adopted under this subchapter apply to the following entities: (1) a health plan; (2) a health care clearinghouse; and (3) a health care provider who transmits any health information in electronic form in connection with a transaction covered by this subchapter" (45 CFR § 160.103; 45th Code of Federal Regulations, Office of the Federal Register, United States of America). When data leave these entities, the imposed safeguards by HIPAA do not apply anymore, resulting in less strict regulations. A further guidance issued by the US Food and Drug Administration (FDA) concerns mHealth apps. It addresses, for example, apps that pose a high risk to the public. Apps only fall within its scope if they transform a mobile phone or any other electronic device into a medical device. As the FDA acknowledges in its guidance document, it does not address a substantial number of health data collectors, such as wellbeing apps; websites, especially patient centered portals like PatientsLikeMe; and social networks, and thus, it excludes most indirect, inferred, and invisible health data, which subsequently are subject to the US Federal Trade Commission guidance, resulting in lower safeguards of potentially highly personal data [17].

In contrast, the European Union General Data Protection Regulation (GDPR) offers a comprehensive framework for any kind of personal data and adds different notions to, for example, health or research data. The GDPR treats health data as a special category of data, which is sensitive by its nature, along with several other types, and the statement is as follows: "Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited." (Article 9, section 1; General Data Protection Regulation, The European Union). Two key points are illustrated by this quotation. First, it is clear that social media data could reveal any of these different sensitive types of data. Second, the term "health data" is not actually mentioned; the phrase used is "data concerning health." This opens the door to indirect and inferred health data falling within the scope of the GDPR.

Processing of health data is prohibited unless exceptions apply, and one of them is the provision of the individual's explicit consent. The collection of consent from the data subject remains one of the most common exceptions that organizations processing health data will be able to rely on when it has been explicitly provided and the purpose for processing the data has been explicitly defined.

Comparing the two legislations (EU and US), it can be said that most US regulations on health data remain at the national or state level, such as the HIPAA in the US, which is tightly attached to parts of the health system, such as hospitals, health insurance companies, and pharmacies. Other important data stemming from social networks and other providers are not

covered by many of these regulations because they are out of their scope. Furthermore, there are no specific regulations other than broad principles, such as the fair information principles, issued by the US trade commission. In contrast, the GDPR offers at least some boundaries, but there are a number of serious questions on how to interpret the approach of the GDPR to the regulation of social media data. Our broadening of the notion of health data implies both that higher safeguards are applied to nonspecial categories of data and that the lawful grounds for processing such kinds of data need to take care of this situation, and imposing implicit consent could be one approach. As today's consent mechanisms are unable to handle this extra burden, there is an urgent need to foster the development of an alternative consent mechanism. Some examples are delineated later in this article.

The issue of different approaches to data protection has been the subject of disagreement in the literature. Cate et al, who are opponents of a more liberal approach, have argued for self-regulation of data protection principles [18], shifting the burden to users. In view of the recent scandals involving major data intensive companies, this view seems rather inappropriate and neglects the ethical responsibilities of companies toward their customers [19,20].

In contrast, authors like McDermont et al advocate data protection as a human right, with its underpinning ethical principles of privacy, transparency, autonomy, and nondiscrimination. These principles are particularly important in light of the increasing use of large amounts of data and algorithmic prediction [21]. This is also highlighted by Wachter et al, who called for a new right of reasonable inferences to close the accountability gap currently posed by "high-risk inferences," especially regarding predictions drawn from big data analytics with low verifiability and thus possible damaging effects on individuals [22].

Extended Notions of Health Data and Consent

Until now, we have assumed that data used in research are primarily generated prospectively (ie, right now or recently) and have excluded data from the more distant past. However, research projects, such as the Time Machine project [23], aim to generate digital copies of vast amounts of past paper-based data. Indeed, digitization of past data is progressing rapidly, for example, in the digital humanities and in the digitalization of large amounts of business and health data [24]. A new notion is thus introduced to the concept of consent. As today's consent is based on the assumption that the future use of data must be regulated, the linking of old or past data in connection with digitalization would also require consent when an identified or identifiable person is concerned. Consent will therefore have to deal with the past, present, and future use of both past and prospective data.

Consent to use pre-existing data is dealt with in a variety of ways. In medicine, research participants are sometimes invited to give "broad consent" to future reprocessing of data, which is subject to review by a research ethics committee. Such consent is necessarily broad because those giving it will often have no idea of the specific research projects that might use their data in the future. Other models require recontacting participants in order to obtain specific consent for each future project. More radical is the concept of "data donation," where people grant access to their data under very few limited conditions, if any [25]. However, all of these models are derived from health care and medical research. In the wider context of social media, financial, and location data, consent is based on the initial agreement to the T&Cs as described above and is mostly of a commercial background. Despite the fact that the data could be used in a myriad of ways, even if not ultimately health related, this consent is often entirely uninformed. This is even the case for data that are currently being generated. Awareness and transparency are even lower in terms of possible uses of past paper-based data that are digitalized or future data that might seem irrelevant now but could yield highly relevant health data when combined with other datasets. Current consent for data sharing is to a large extent blind owing to its broad nature. Consent systems actually need to look far back and far forward, as well as in close detail at the present. In essence, consent must be capable of time travel, just as data are capable of time travel.

Toward Comprehensive Consent in a Hyperconnected World

Several scholars have pointed out that traditional models of obtaining consent have reached their limits in today's highly data-driven and data-intensive research, and this articulates the need for new forms for obtaining participant consent [26,27]. Concretely, the traditional form of obtaining consent by individually informing participants about their rights and protections is practically impossible in such environments [28-30] owing to the sheer scale and challenges associated with such an endeavor. Several scholars have proposed possible ways to tackle this issue, ranging from information technology-based systems like dynamic consent [31], which offer a better way to inform and maintain a relationship between researchers and participants, to stewardship-based solutions, such as those where a community-based approach assures data governance [32], and radical solutions like data donation [25]. However, most of today's research projects simply use a digitized version of traditional consent procedures.

As we have already pointed out, past data will become increasingly health relevant. This applies particularly to public and commercial research where it is expected that increasingly more research will be data driven and data will stem from many disparate data sources, including commercial sources. There is thus an urgent need for a better approach of informing customers in a truly informative way. However, as stated above [16], the current approaches are ineffective as T&Cs are too long and written in a too complicated way, undermining what is at the heart of genuinely informed consent, namely the prevention and hold up of basic ethical principles [17,18]. A possible way to tackle this is to introduce harmonized T&Cs, which would need stronger government interventions. Further possibilities are to move toward comic-based consent as developed by

Brunschwig for contract law [33] and to implement "nutrition label-like" consent [34] for T&Cs.

The four distinct categories of direct, indirect, inferred, and invisible health data mentioned earlier in this article may each require a differentiated consent solution, and the past, present, and future aspects of consent for the use of big data complicate the situation further.

Direct health data are most easily governed, although they have their own set of challenges. A specific consent system may seem simple but can impose substantial limitations on researchers and prove very burdensome for patients. Broad consent poses its own set of problems, particularly in terms of future data linkage. Even with specific consent to use an individual's data in a particular project, researchers might also want broad consent to access a participant's entire medical history and link it to medical records to facilitate follow-up, meaning that consent is provided for health data generated in not only the project itself (the present), but also the years or decades before it (the past) and the many years to come (the future). The growing discussion around data donation illustrates the ethical and legal challenges related to providing consent to ongoing use beyond death for an entire lifetime of genetic and nongenetic health data, some of which may have implications for relatives. In fact, most of the current data protection frameworks, such as GDPR, neglect data donation, and use of data after death is not within their scope (Recital 27 EU GDPR). At least until the point of death, a dynamic consent system seem to be a promising means of controlling different users' access to past and present medical data and for controlling data linkage with other studies. Another issue complicating the use of direct data concerns mHealth data (ie, data that are gathered when using mHealth solutions). Such apps are mostly overseen by national authorities, and, for instance, in the United States, they need approval from the FDA. Data gathered by those apps need higher protection by law than US regulations and GDPR provide at present. In particular, the issue of consent is rather unregulated, and what could be seen as possible consent (the acceptance of the T&Cs) does not meet the high standards as imposed by traditional consent in research.

Indirect health data, such as exercise, social media, and movement data concerning an individual, are not currently regulated within the health data model in the United States. In the European Union, the GDPR covers such data, but they are not regarded as distinct health data and thus benefit from less protection, as is the case for inferred data. As stated above, consent is usually given via the T&Cs of relevant apps. This must change in the future. Either T&Cs must become much more user friendly and accessible or an entirely different consent model and system more akin to direct health data governance will have to be adopted.

Inferred data are particularly problematic in terms of consent, as they are the result of the combination of two or more datasets of one individual. Consent may have been given to the processing of each of these data points (or sets) but not to their combined processing, which can yield more revealing data not anticipated at the time consent was provided. One way to approach this problem is to make it clear at the point of consent to use direct and indirect health data that their combination with other datasets is a real possibility despite goodwill and efforts on the part of researchers, companies, and other users to prevent it. With further technical developments, it might even be possible to send an alert to a dynamic consent portal for each new instance of combining data points of the individual, enabling tighter and more finely grained control of inferred data. This would ensure that consent is provided when new inferences are made. It is important to bear in mind in the discussion of inferred health data that some data, particularly genetic data, affect not only the data subject but also family members, which is also true for social media data when parents share information of their close relatives and children. Some individuals might be very happy to share all types of data, but this can have ramifications for close relatives, particularly identical twins, in the case of genetic data. In such cases, some form of collective consent may be required.

By its very nature, consent to use invisible health data cannot currently be provided, as we are still blind to the very nature of such data and their potential relevance to health. However, consent can be "future proofed" to a limited extent with careful legislation and regulation. By adopting a similar level of oversight for all types of data concerning a person (the GDPR is a step in this direction), safeguards will be in place once it emerges that seemingly entirely innocuous data can be used by artificial intelligence technologies to yield health findings. Once this transpires, alerts to dynamic consent systems will be a sensible precaution.

As the future use of data cannot be foreseen at present, alert mechanisms play a particularly important role, especially for inferred health data and invisible health data. Given our thesis that any data could turn out to be relevant to health, alerts might well be essential to ensure that people are kept informed when their data are put to a novel use with health implications. Whether citizens would or could have any right to stop the processing of such "new" health data is a difficult question that is outside the scope of this paper, but informing them seems to be a basic ethical requirement. If, as we suggest, all data will become health relevant, it might be impossible or, at least, very impractical for people to stop the processing of all data relevant to their health. Laws, such as the GDPR, will have to keep pace with developments in the conception of health data, as imposing current GDPR standards on all data that might be relevant for health in the future could have great implications for the processing of data, particularly in research. Table 2 summarizes the different categories of health-related data and how consent to use past, present, and future data could be approached.

Table 2. Consent based on different health data types related to their temporal origin.

Data type	Past data	Present data	Future data
Direct health data	Specific/broad consent	Direct/specific/broad consent	Direct consent
Indirect health data	Direct consent/terms and conditions	Direct consent/terms and conditions	Direct consent
Inferred health data	N/A[a]	Alerts to dynamic consent systems	Alerts to dynamic consent systems
Invisible health data	N/A	N/A	Alerts to dynamic consent systems

[a]N/A: not applicable.

Conclusion

With each passing day, billions of gigabytes of direct, indirect, and inferred health data are being recorded, with massive implications for privacy and harm prevention if adequate consent mechanisms for their use are not in place. The possibility of invisible health data complicates the situation further. If all our data will be health data one day, we need to start treating consent to data use with the respect that it deserves. Currently, most data collectors are gaining access to vast amounts of behavioral and health data effectively for free, without having to comply with any safeguards. Broadening the notion of health data, as we have suggested, would cause companies to give more thought to ethical acquisition and processing of data. However, broadening the notion of health data could have an adverse effect on research if it results in excessively burdensome regulations.

Our argument that all data are health data is primarily ethical, but it could have important legal ramifications. In jurisdictions where health data can only be processed with consent, widening the scope of health data in this way would vastly increase the burden on, for example, private companies who process indirect and inferred health data. This might be difficult but ethically appropriate, and the development of more modern and dynamic consent mechanisms could facilitate this shift. Alternatively, legislation could limit the legal scope of health data to direct health data, leaving soft laws and guidelines to regulate other categories of health data.

Acknowledgments

This research was funded by the Swiss National Science Foundation Project (number: 167211).

Conflicts of Interest

None declared.

References

1. Ienca M, Ferretti A, Hurst S, Puhan M, Lovis C, Vayena E. Considerations for ethics review of big data health research: A scoping review. PLoS One 2018;13(10):e0204937 [FREE Full text] [doi: 10.1371/journal.pone.0204937] [Medline: 30308031]
2. Reece A, Danforth C. Instagram photos reveal predictive markers of depression. EPJ Data Sci 2017 Aug 8;6(1) [FREE Full text] [doi: 10.1140/epjds/s13688-017-0110-z]
3. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. J Med Internet Res 2015 Jul 15;17(7):e175 [FREE Full text] [doi: 10.2196/jmir.4273] [Medline: 26180009]
4. Cao H, Lin M. Mining smartphone data for app usage prediction and recommendations: A survey. Pervasive and Mobile Computing 2017 Jun;37:1-22. [doi: 10.1016/j.pmcj.2017.01.007]
5. Do TM, Gatica-Perez D. Where and what: Using smartphones to predict next locations and applications in daily life. Pervasive and Mobile Computing 2014 Jun;12:79-91. [doi: 10.1016/j.pmcj.2013.03.006]
6. Kaplan B. How Should Health Data Be Used? Camb Q Healthc Ethics 2016 Mar 09;25(2):312-329. [doi: 10.1017/s0963180115000614]
7. Cate F, Mayer-Schonberger V. Notice and consent in a world of Big Data. International Data Privacy Law 2013 May 20;3(2):67-73 [FREE Full text] [doi: 10.1093/idpl/ipt005]
8. UK Children's Commissioner. 2018. Simplified social media terms and conditions for Facebook, Instagram, Snapchat, YouTube and WhatsApp URL: https://www.childrenscommissioner.gov.uk/publication/simplified-social-media-terms-and-conditions-for-facebook-instagram-snapchat-youtube-and-whatsapp/ [accessed 2019-10-24]
9. The European Parliament and The Council of The European Union. Official Journal of the European Union. General Data Protection Regulation URL: https://eur-lex.europa.eu/ [accessed 2019-10-24]
10. Jain SH, Powers BW, Hawkins JB, Brownstein JS. The digital phenotype. Nat Biotechnol 2015 May;33(5):462-463. [doi: 10.1038/nbt.3223] [Medline: 25965751]

11. Kosinski M, Stillwell D, Graepel T. Private traits and attributes are predictable from digital records of human behavior. Proc Natl Acad Sci U S A 2013 Apr 09;110(15):5802-5805 [FREE Full text] [doi: 10.1073/pnas.1218772110] [Medline: 23479631]
12. Vayena E, Blasimme A. Biomedical Big Data: New Models of Control Over Access, Use and Governance. J Bioeth Inq 2017 Dec 5;14(4):501-513 [FREE Full text] [doi: 10.1007/s11673-017-9809-6] [Medline: 28983835]
13. Vayena E, Blasimme A. Health Research with Big Data: Time for Systemic Oversight. J Law Med Ethics 2018 Mar;46(1):119-129 [FREE Full text] [doi: 10.1177/1073110518766026] [Medline: 30034208]
14. Mayer M, Fernández-Luque L, Leis A. Big Data For Health Through Social Media. In: Participatory Health Through Social Media. Cambridge, Massachusetts: Academic Press; 2016:67-82.
15. Harari G, Lane N, Wang R, Crosier B, Campbell A, Gosling S. Using Smartphones to Collect Behavioral Data in Psychological Science: Opportunities, Practical Considerations, and Challenges. Perspect Psychol Sci 2016 Nov;11(6):838-854 [FREE Full text] [doi: 10.1177/1745691616650285] [Medline: 27899727]
16. Jain P, Gyanchandani M, Khare N. Big data privacy: a technological perspective and review. J Big Data 2016 Nov 26;3(1):a [FREE Full text] [doi: 10.1186/s40537-016-0059-y]
17. Federal Trade Commission. 2016. Mobile Health App Developers: FTC Best Practices URL: https://www.ftc.gov/tips-advice/business-center/guidance/mobile-health-app-developers-ftc-best-practices [accessed 2019-10-31]
18. Cate FH, Cullen P, Mayer-Schonberger V. Data Protection Principles for the 21st Century. In: Maurer Faculty. Redmond, WA: Microsoft Corporaton; 2013.
19. Schneble CO, Elger BS, Shaw D. The Cambridge Analytica affair and Internet-mediated research. EMBO Rep 2018 Aug;19(8) [FREE Full text] [doi: 10.15252/embr.201846579] [Medline: 29967224]
20. Schneble CO, Elger BS, Shaw DM. Google's Project Nightingale highlights the necessity of data science ethics review. EMBO Mol Med 2020 Mar 06;12(3):e12053 [FREE Full text] [doi: 10.15252/emmm.202012053] [Medline: 32064790]
21. McDermott Y. Conceptualising the right to data protection in an era of Big Data. Big Data & Society 2017 Jan;4(1):205395171668699. [doi: 10.1177/2053951716686994]
22. Wachter S, Mittelstadt B. A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI. Columbia Bus Law Rev 2019 Apr 1:1-130 [FREE Full text] [doi: 10.31228/osf.io/mu2kf]
23. Time Machine. Big Data of the Past for the Future of Europe URL: https://www.timemachine.eu/ [accessed 2019-10-24]
24. Jones L, Nevell R. Plagued by doubt and viral misinformation: the need for evidence-based use of historical disease images. Lancet Infect Dis 2016 Oct;16(10):e235-e240. [doi: 10.1016/S1473-3099(16)30119-0] [Medline: 27522232]
25. Shaw DM. Defining Data Donation After Death: Metadata, Families, Directives, Guardians and the Route to Big Consent. In: Krutzinna J, Floridi L, editors. The Ethics of Medical Data Donation. Cham: Springer; 2019.
26. Mostert M, Bredenoord A, Biesaart M, van Delden JJ. Big Data in medical research and EU data protection law: challenges to the consent or anonymise approach. Eur J Hum Genet 2016 Jul;24(7):956-960 [FREE Full text] [doi: 10.1038/ejhg.2015.239] [Medline: 26554881]
27. Ioannidis J. Informed consent, big data, and the oxymoron of research that is not research. Am J Bioeth 2013;13(4):40-42 [FREE Full text] [doi: 10.1080/15265161.2013.768864] [Medline: 23514395]
28. Christen M, Domingo-Ferrer J, Draganski B, Spranger T, Walter H. On the Compatibility of Big Data Driven Research and Informed Consent: The Example of the Human Brain Project. In: Mittelstadt B, Floridi L, editors. The Ethics of Biomedical Big Data. Cham: Springer; 2016:199-218.
29. Kaye J, Whitley EA, Lund D, Morrison M, Teare H, Melham K. Dynamic consent: a patient interface for twenty-first century research networks. Eur J Hum Genet 2015 Feb;23(2):141-146 [FREE Full text] [doi: 10.1038/ejhg.2014.71] [Medline: 24801761]
30. Jacobs B, Popma J. Medical research, Big Data and the need for privacy by design. Big Data & Society 2019 Jan 18;6(1):205395171882435 [FREE Full text] [doi: 10.1177/2053951718824352]
31. Williams H, Spencer K, Sanders C, Lund D, Whitley EA, Kaye J, et al. Dynamic consent: a possible solution to improve patient confidence and trust in how electronic patient records are used in medical research. JMIR Med Inform 2015 Jan 13;3(1):e3 [FREE Full text] [doi: 10.2196/medinform.3525] [Medline: 25586934]
32. Blasimme A, Vayena E, Hafen E. Democratizing Health Research Through Data Cooperatives. Philos. Technol 2018 Jun 19;31(3):473-479 [FREE Full text] [doi: 10.1007/s13347-018-0320-8]
33. Brunschwig C. Visualisierung von Rechtsnormen: Legal Design. Zürich: Schulthess; 2001.
34. Kelley P, Bresee J, Cranor L, Reeder R. A "nutrition label" for privacy. In: Proceedings of the 5th Symposium on Usable Privacy and Security - SOUPS '09. New York, New York, USA: ACM Press; 2009 Presented at: SOUPS '09: Symposium on Usable Privacy and Security; July 2009; Mountain View, California, USA p. 1-12 URL: http://portal.acm.org/citation.cfm?doid=1572532.1572538 [doi: 10.1145/1572532.1572538]

Abbreviations

FDA: Food and Drug Administration
GDPR: General Data Protection Regulation

HIPAA: Health Insurance Portability and Accountability Act
mHealth: mobile health
T&Cs: terms and conditions

Edited by G Eysenbach; submitted 04.11.19; peer-reviewed by C Skirrow, E Whitley, D Carvalho; comments to author 23.12.19; revised version received 17.02.20; accepted 22.03.20; published 28.05.20.

Please cite as:
Schneble CO, Elger BS, Shaw DM
All Our Data Will Be Health Data One Day: The Need for Universal Data Protection and Comprehensive Consent
J Med Internet Res 2020;22(5):e16879
URL: http://www.jmir.org/2020/5/e16879/
doi:10.2196/16879
PMID:32463372

Tutorial

Using Intervention Mapping to Develop a Digital Self-Management Program for People With Type 2 Diabetes: Tutorial on MyDESMOND

Michelle Hadjiconstantinou[1], BSc, MSc, PhD, CPsychol; Sally Schreder[2], MSc; Christopher Brough[2], BSc; Alison Northern[2]; Bernie Stribling[2], MBA; Kamlesh Khunti[1], MBChB, PhD, MD, FRCGP, DCH, DRCOG; Melanie J Davies[1], MB, ChB, MD, FRCP, FRCGP

[1]Diabetes Research Centre, University of Leicester, Leicester, United Kingdom

[2]Leicester Diabetes Centre, NHS Trust, University Hospitals of Leicester, Leicester, United Kingdom

Corresponding Author:
Sally Schreder, MSc
Leicester Diabetes Centre
NHS Trust
University Hospitals of Leicester
Gwendolen Road
Leicester
United Kingdom
Phone: 44 116 258 4320
Email: sally.schreder@uhl-tr.nhs.uk

Abstract

Digital health interventions (DHIs) are increasingly becoming integrated into diabetes self-management to improve behavior. Despite DHIs becoming available to people with chronic conditions, the development strategies and processes undertaken are often not well described. With theoretical frameworks available in current literature, it is vital that DHIs follow a shared language and communicate a robust development process in a comprehensive way. This paper aims to bring a unique perspective to digital development, as it describes the systematic process of developing a digital self-management program for people with type 2 diabetes, MyDESMOND. We provide a step-by-step guide, based on the intervention mapping (IM) framework to illustrate the process of adapting an existing face-to-face self-management program (diabetes education and self- management for ongoing and newly diagnosed, DESMOND) and translating it to a digital platform (MyDESMOND). Overall, this paper describes the 4 IM steps that were followed to develop MyDESMOND—step 1 to establish a planning group and a patient and public involvement group to describe the context of the intervention and program goals, step 2 to identify objectives and determinants at early design stages to maintain a focus on the strategies adopted, step 3 to generate the program components underpinned by appropriate psychological theories and models, and step 4 to develop the program content and describe the iterative process of refining the content and format of the digital program for implementation. This paper concludes with a number of key learnings collated throughout our development process, which we hope other researchers may find useful when developing DHIs for chronic conditions.

(J Med Internet Res 2020;22(5):e17316) doi:10.2196/17316

KEYWORDS

diabetes mellitus, type 2; technology; self-management

Introduction

Background

Type 2 diabetes (T2D) is a progressive lifelong condition affecting over 422 million people worldwide [1] and is associated with a number of severe macrovascular complications (damage to large blood vessels leading to coronary heart disease and stroke) and microvascular complications (damage to small blood vessels leading to kidney disease, blindness, and amputation). Diabetes self-management has been shown to reduce the risk of these related complications, improve one's well-being, and reduce the UK National Health Service's (NHS) health costs [2].

Self-management education can be delivered in several forms; one of these forms is via structured education programs. Structured education programs are nationally and internationally recommended for all people with T2D and meet nationally agreed criteria, including following an evidence-based curriculum that is theory-driven, evidence-based, and resource-effective, carrying out quality assurance of teaching standards and delivering regular audits [3]. Examples of such self-management education programs are diabetes education and self-management for ongoing and newly diagnosed (DESMOND) [4,5] and X-PERT (Diabetes education programme for people with T2D) [6] in the United Kingdom, Rethink Organization to iMprove Education and Outcomes [7] in Italy, and Structured Intensive Diabetes Education Program in South Korea [8]. These programs are available for people with T2D, both newly diagnosed and established [9].

In 2008, the National Institute of Health and Clinical Excellence published guidelines to recommend the availability of group-based structured education programs as a key priority for diabetes management [10]. Evidence has shown that group-based structured education programs are essential for effective diabetes management, aiming to improve people's knowledge, skills, and confidence, enabling them to take control of their own condition. Effective education programs have also been shown to improve biomedical (weight and glucose levels), behavioral (ie, physical activity), and psychosocial outcomes (ie, diabetes-related distress and self-efficacy) [9]. Various barriers, however, appear to be associated with low uptake and attendance at structured education programs [11,12], with suboptimal access to programs commonly noted [13]. Attendance and accessibility for patients may also be compromised by health care professionals' insufficient advocacy for structured education programs, as well as limited infrastructure and capacity to offer face-to-face programs [14].

Although referrals to structured education have increased over the last 3 years, attendance at such programs have decreased, and attendance rates are as low as they were in 2009 [15]. This is why a recent call by NHS England in the UK advised that people with diabetes are offered face-to-face programs, and those who cannot attend must, then, be offered a digital alternative [16].

Digital health interventions (DHIs, which are any digital-based intervention, such as smartphones, websites, text messaging) [17] have become a means of delivering care and offering support to facilitate behavior change in individuals with long-term conditions. Mounting evidence indicates that the use of digital technology for diabetes self-management is an effective way to support people to live well with their diabetes [18,19], with programs showing improvements mainly in glycemic control [18,20].

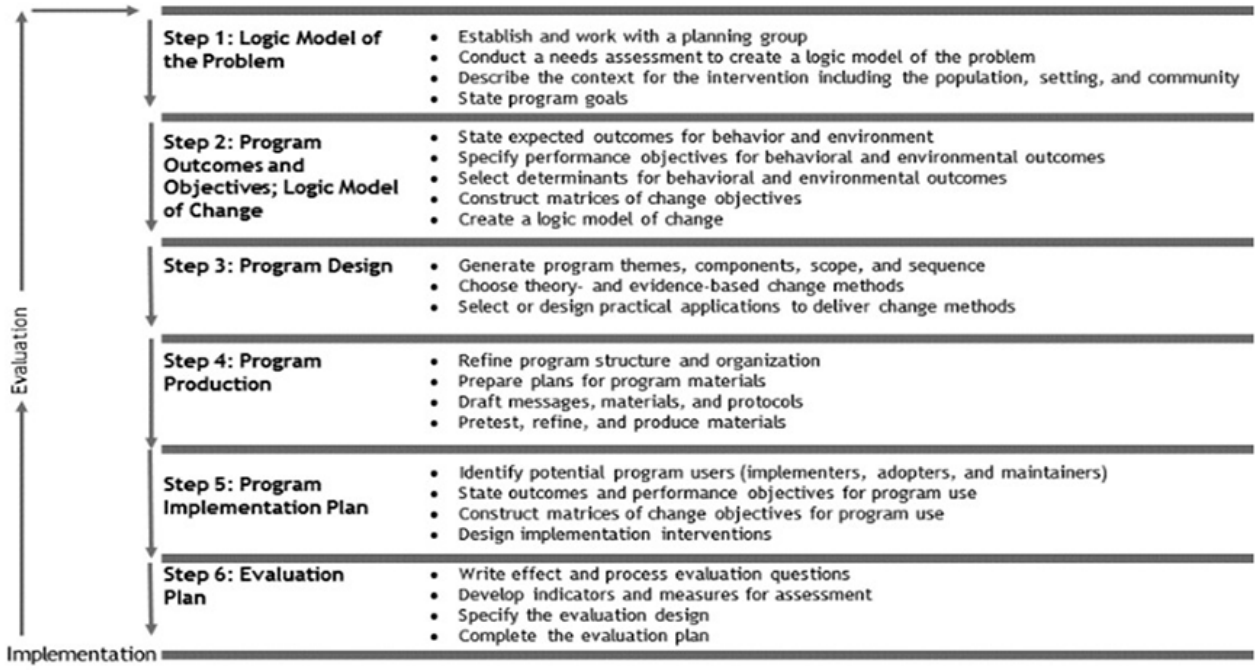
DHIs in other chronic conditions have been shown to be cost-effective, with a recent cost-effectiveness study depicting the benefits and cost-effectiveness of digital-based psychological interventions for reducing behavioral stress management and anxiety [21,22]. Information derived from DHIs can be kept up to date more easily with local and national guidelines and recommendations. DHIs are neither geographically nor time restrained. Indeed, barriers associated with face-to-face structured education programs, such as time and transport, are often easier to address with digital interventions, as users have the opportunity to access Web-based programs and read the material at their own time and pace.

Appropriate steps are essential when developing evidence- and theory-based digital programs to ensure they are implemented effectively in the real world [23]. Despite DHIs becoming more and more available to people with diabetes, the development strategies and processes undertaken can be somewhat unclear. With theoretical frameworks available in the current literature, such as Bartholomew's intervention mapping (IM) framework [24], it is important that DHIs follow a shared language and communicate a robust development process appropriately and systematically [25].

Intervention Mapping

IM, a 6-step framework, is often applied to guide behavior change interventions and health education development. This evidence-based approach, although presented as a series of 6 steps, follows a more iterative process that bridges the gap between theory and practice. With this in mind, we aimed to develop our digital program based on a systematic and robust process such as IM, to ensure that our evidence-based program can be implemented in real-world settings. The development process for MyDESMOND was guided by the first 4 steps of IM, as illustrated in Figure 1: (1) conduct a needs assessment, (2) specify program outcomes and objectives, (3) design program and apply theory, and (4) refine program development. The last 2 steps (5) and (6) of the IM framework for MyDESMOND (adoption, implementation, and evaluation plan) are currently being evaluated.

Figure 1. Intervention mapping framework, Bartholomew.



Aim

This paper describes the systematic developmental process that we undertook to adapt the content and evidence of an existing face-to-face structured self-management education program (DESMOND) for T2D and translate it into a digital self-management program (MyDESMOND). This process is presented based on the IM framework [24].

Methods

Step 1: Needs Assessment

Establish and Work With a Planning Group

A multidisciplinary team (psychologist, dietitian, diabetes specialist nurses, and experts in physical activity and digital health) was brought together to provide expertise in the psychosocial and clinical aspects of T2D, digital health, and behavior change research. Throughout the design process, our planning team worked alongside a Web-designer to ensure that the logistics and digital platform would be appropriate for a UK-based program.

Patient and public involvement was essential for our development process [26]. Throughout our process, we worked closely with people with T2D who provided feedback on the content: format and usability of MyDESMOND (see Step 4 Program Development for further details).

Conduct Mixed Methods Needs Assessment

We conducted a preliminary scoping search, and together with the findings of our existing research studies, we aimed to collate key learning points to explore and identify the key program goal and key behavioral determinants of MyDESMOND.

Scoping Literature

Self-management involves 3 key tasks (medical, role, and emotional) and consists of core self-management skills, including problem-solving, decision making, resource utilization, and action planning [27]. Digital self-management interventions over recent years have provided users with the knowledge and skills required to make a lifestyle change and improve their self-care. Enabling people with diabetes to better understand their conditions has been shown to improve coping strategies and reduce health-related anxiety and depression [18].

In addition to knowledge and improved coping skills, social support has also been linked as a major predictor of, and positive influence on diabetes self-management. A systematic review, which explored the role of social support on biomedical and psychosocial outcomes, as well as behavior change, reported a positive impact on the aforementioned outcomes [28]. In addition, recent diabetes literature has also illustrated a strong association between social support and improved diabetes self-management and self-efficacy [29].

A qualitative evaluation of patient and health care professional (HCP) views on barriers to self-management and multimorbidity (including diabetes) identified 3 main factors that influence self-management: (1) capacity (access and availability of resources, time, and knowledge), (2) responsibility (the degree to which patients and HCPs agree about self-management), and (3) motivation (willingness to take-up types of self-management practices) [30].

Systematic Reviews

We conducted focused synthesis of work in 2 areas: (1) a systematic review and meta-analysis to critically appraise and quantify the evidence on the effect of digital interventions that aim to improve well-being in people with T2D [31], and (2) a more recent systematic review of systematic reviews to explore

and synthesize evidence around the applications of digital patient education in T2D and cardiovascular disease [32].

Systematic Review and Meta-Analysis

The systematic review and meta-analysis [31] identified preliminary evidence suggesting that digital interventions with significant well-being outcomes (in this case, depression or diabetes-related distress outcomes) shared common characteristics: (1) all interventions provided professional-led support, (2) interventions consisted of both asynchronous and synchronous communication, and (3) intervention duration was between 2 and 6 months. This review also identified that the common behavior change techniques (BCTs) adopted by effective digital interventions were *information provision* and *monitoring*.

The aim of this review was to examine if digital-based interventions improve well-being outcomes such as depression and diabetes-related distress; however, no significant improvements were found following a meta-analysis in depression (P=.15) or distress (P=.43). Nevertheless, the review did take into consideration that the number of studies included in the meta-analysis was low, and thus, conclusions on the meta-analysis were considered with caution.

On the basis of the findings above, we considered the following elements in our development work: (1) to have a professional-led, rather than peer-led program, (2) to include some form of *monitoring* element and feedback to users, and (3) to include both direct and indirect communication, in the form of a chat forum or messages.

Systematic Review of Systematic Reviews

This systematic review of the reviews [32] explored and synthesized the scope of digital patient education in T2D on patient outcomes. Our findings indicated that biological, behavioral, and psychological outcomes, as well as knowledge and self-efficacy for people with T2D, had shown examples of benefits from digital patient education. In addition, outcomes that most consistently showed improvement with the intervention (irrespective of effect size) were weight, physical activity, knowledge, social support, and quality of life. These findings suggest that digital patient education has a wide range of benefits for people with T2D. Therefore, we considered these findings in our development work and ensured that our program incorporated biological, behavioral, and psychological outcomes. Self-efficacy was also another crucial outcome, which we felt was important for behavior change (see additional information in theories, Step 3).

Our Current Face-to-Face Self-Management Program (DESMOND)

Our face-to-face self-management program [33] was developed for people with newly diagnosed T2D with the aim that participants would focus on lifestyle behaviors (such as food choices and physical activity), consider their personal risk factors, and choose a specific achievable goal to work on. This program is delivered for 6 hours in the community by 2 trained educators and is based on psychological theories of learning: Leventhal common sense theory [34], dual-process theory [35], and social learning theory [36], and the philosophy of the program was grounded on patient empowerment [37]. On the basis of the evaluation of DESMOND [33], this group self-management program showed greater improvements in weight loss and positive improvements in illness beliefs. Our aim was to adopt key elements of DESMOND that we considered crucial to the effectiveness of this program, including group dynamic and peer support, interactive activities, and setting a personal goal to help with behavior change. To ensure that elements were retained in the adapted format and remained suitable for a digital format, we found it was essential to revisit and underpin theories to our digital program that focused on the process of change. Thus, with the aforementioned key elements in mind, we aimed to provide a virtual environment whereby Web users would have the opportunity to share thoughts and experiences in the form of a chat forum and retain the peer support element. Interactive activities were developed to ensure that information was not provided in a didactic way, but instead was provided in a way whereby Web users could reflect on answers and aid their confidence. Similar to DESMOND, MyDESMOND users were able to assess which behavior change is important to them and tailor a personal action plan suitable to their environment and needs. Users would then be able to revisit their action plan to assess their progress. From preliminary feedback, participants expressed satisfaction with the content of the program (lifestyle behaviors) and reported they had gained knowledge and learned new skills in healthy living.

Our Digital-Based Self-Management Program

BabySteps

In 2017, we developed our first digital-based self-management program on a responsive digital platform to promote physical activity in women with a history of gestational diabetes (BabySteps) [38]. This digital program consisted of functions such as information material, step challenges, chat forums, and health trackers, and these functions were based on gamification and BCTs, including problem-solving, coping skills, and action planning. With the BabySteps trial coming to an end, we carried out feedback sessions with 12 participants to explore user experience and satisfaction. Preliminary findings have helped structure MyDESMOND in several ways: (1) DHIs are needed for chronic conditions, (2) to adopt a new behavior, people need to learn and practice self-care skills, (3) the navigation on apps plays a crucial role in digital uptake, (4) people engage with digital programs if it comes from a trustworthy source, and (5) peer support is essential in DHIs.

MyWellbeing

It was important to consider that BabySteps was a prevention program with a focus on physical activity (ie, increase steps and reduce sedentary behavior). We therefore supplemented our needs assessment step with further work conducted by our lead author [39] to develop a digital emotional support program to improve diabetes-related distress in T2D (MyWellbeing). This additional work aimed to complement the evidence behind MyDESMOND by incorporating the element of emotional management. This work adopted a person-centered approach [40] and gathered evidence from exploratory mixed methods research work (which will be published elsewhere). Most

importantly, this work was based on ongoing patient and public involvement (PPI) engagement to support the development of this digital-based program. Key learnings from the PPI groups (which consisted of people with T2D and carers) were as follows: (1) incorporate links to useful external resources, (2) design a program with a positive language and motivational content, (3) incorporate *ask the expert* function, (4) use short videos to provide information, (5) include quizzes as a form of interactive learning, and (6) consider a *frequently asked question* page to enhance engagement.

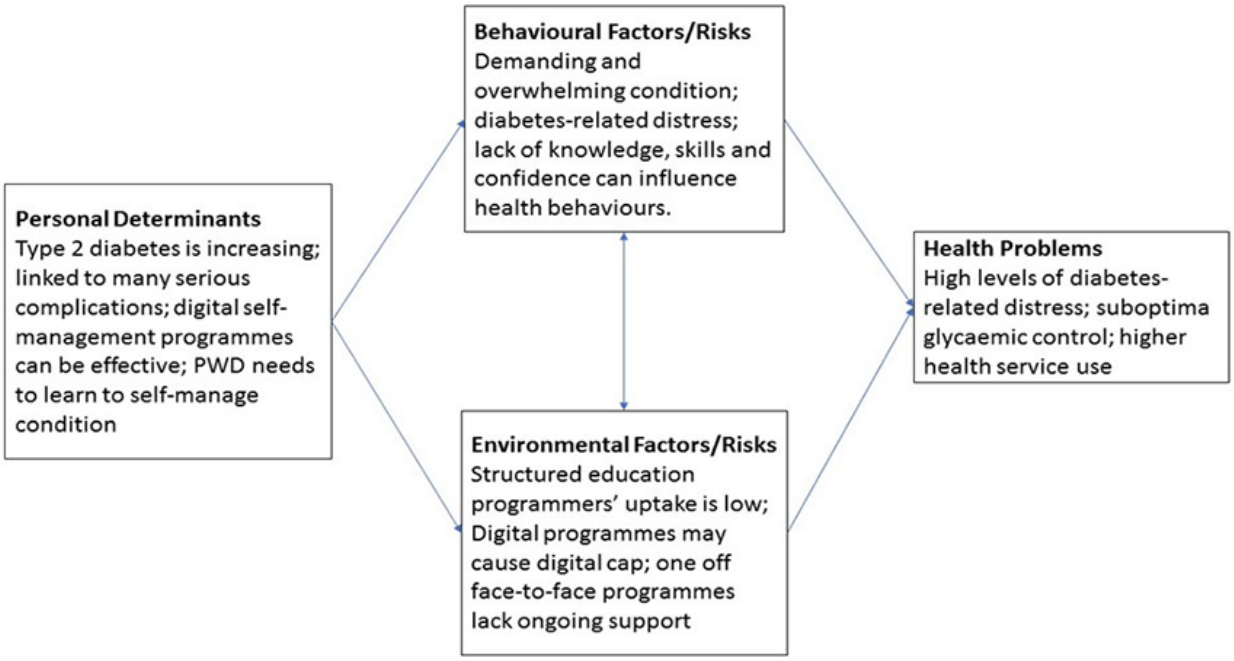
Identify Program Goals

Overall, the outcomes of the needs assessment provided the evidence for our program goals and defined the purpose of our digital program, which was to provide a digital self-management program to people with T2D to improve medical, behavioral, and emotional self-management. This was developed with the aim to improve and maintain self-management behaviors, improve knowledge, promote self-care skills, increase self-efficacy, and improve well-being.

Create a Logic Model of the Problem

On the basis of our scoping literature and mixed methods studies (as part of step 1), we produced a logic model of the problem, illustrating the link between behavioral determinants and the behavioral and environmental factors, which lead to the targeted health problems (Figure 2).

Figure 2. Logic model of the problem, step 1 IM.



Step 2: Behavioral Determinants and Performance Objectives

Select Behavioral Determinants and Performance Objectives

This step involved developing the specific change objectives and identifying specific behaviors that determine behavior change for this target population. Informed by the findings from our needs assessment, our planning group identified 5 key behavioral determinants and 5 performance objectives (objectives to be accomplished by the individual to achieve the program goals), which we anticipated would influence the behavioral outcome (Table 1). In detail, behavioral determinants were identified as follows: (1) knowledge, (2) self-efficacy, (3) social support, (4) intention, and (5) behavioral skills. Performance objectives were defined as follows: (1) learn about T2D, (2) be motivated to self-manage, (3) increase access to resource, (4) engage with the digital program, and (5) learn about self-care behaviors. The behavioral determinants listed in the matrix table (Table 1), derived from the Capability, Opportunity, Motivation, Behaviour (COM-B) model (which we explain in the following section), focus on factors that we identified as important in this target population. In this step, we elaborate on the strategies required to ensure that performance objectives are met, and subsequently, determinants are addressed (Table 1).

Table 1. Matrix of objectives.

Behavioral determinants	Performance objectives	Strategies
Knowledge	Learn about T2D[a]	Users gain knowledge; Information on healthy living, stress management, medication, physical activity, and diet
Self-efficacy	Be motivated to self-manage	Users formulate their own goal setting and action plan via patient testimonials and *expert* videos; Users are exposed to modeled behavior
Social support	Increase access to resource	Users become aware of local support; Users receive support not only from HCPs[b] but also from peers
Intention	Engage with the digital program	Users engage with program functions tailored to their needs; intention to change behavior
Behavioral skills	Learn about self-care behaviors	Users practice self-care behaviors

[a]T2D: type 2 diabetes.

[b]HCP: health care practitioner.

Step 3: Theory-Based Intervention Methods and Practical Applications

Identify Theory-Based Methods and Select Practical Applications

In a complex environment, dealing with influences on an individual, social, and cultural level, we aimed to adapt and develop a digital program that was theory-based to break down targeted health behaviors. Theory-based interventions are more effective than non–theory-based interventions and can provide a clear understanding of what works in terms of adapting interventions across populations, behaviors, or contexts [41]. Thus, we adapted the following theoretical frameworks and models to the development of our digital program.

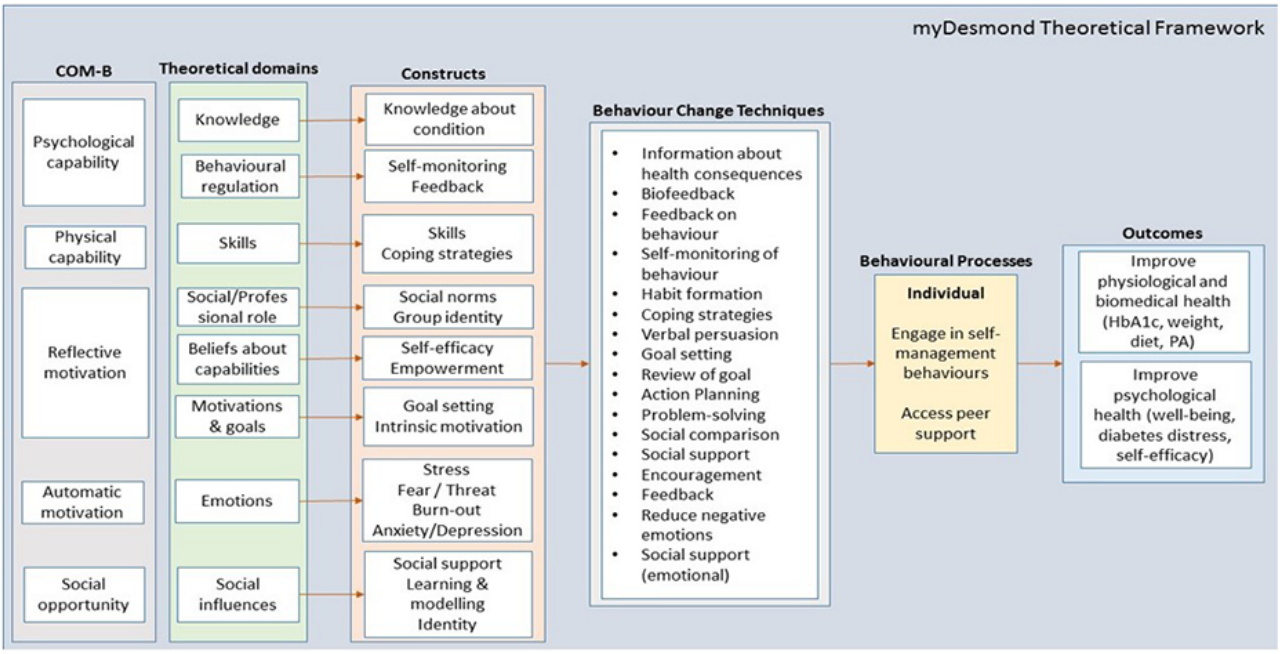
Corbin and Strauss Model on Self-Management Framework

A key conceptual framework of self-management was developed by Corbin and Strauss. Self-management programs (face-to-face or digital) often include at least one of the three main overarching areas: cognitive (such as knowledge, understanding, and self-efficacy), behavioral (such as lifestyle changes and managing medication effectively), and emotional (such as distress, shame, and depression) [21]. On the basis of the Corbin and Strauss model, we aimed to adapt and develop an effective program that considers all three elements for self-management equally: cognitive, emotional, and behavioral [42].

Capability, Opportunity, Motivation, Behaviour Model and the Taxonomy of Behavior Change Techniques

We applied the COM-B model as it provides a common language for the reporting of behavior change interventions [43]. Atkins et al argued the importance of including behavioral determinants and BCTs in the development process of interventions, as these may influence and change health behavior [44,45]. Interventions targeting behavior determinants and behavior change have also been shown to be more effective [46]. Informed by Michie's Behavior Change Wheel (a systematic guide to designing interventions) [43], we populated a theoretical framework for MyDESMOND to identify the relevant mechanisms of action, component constructs, and BCTs, all aimed to inform the individual and organizational processes and health outcomes (Figure 3). Our mechanisms of action, which are processes that influence behavior, included key constructs: knowledge (an awareness of the existence of a condition), behavioral regulation (self-efficacy and feedback), skills (an ability acquired through practice), beliefs about capabilities (belief and acceptance of one's ability to perform), social and professional role and identity (one's displayed behaviors and qualities within a social setting), emotions (one's reaction and attempt to deal with a significant event), and social influences (interactions that influence one's thoughts, feelings, emotions). Table 2 illustrates the relationship between the mechanisms of action, BCTs, and the practical application to implement change methods in MyDESMOND, an important factor for program development.

Figure 3. MyDESMOND theoretical framework.



Health Action Process Approach Model

Despite the application of the behavior change taxonomy, it was still essential to adopt a model that helped explore the process of change and helped explain how individuals decide to engage in healthy behavior. Known as a *process* model, the Health Action Process Approach (HAPA) model informed the development of our digital program. Intention alone does not always change behavior [47], thus with the guidance of action planning and coping planning (both incorporated in our digital program), we aimed to build a program that focused on two processes: goal setting and goal pursuit. With self-efficacy as an additional HAPA construct, we know that people with higher self-efficacy are more likely to sustain the behavior; thus, we ensured that MyDESMOND also incorporated elements for this construct.

Social Support Theory

Social support plays an integral part in the management of diabetes and is often associated with health behaviors such as physical activity. Thus, it was deemed appropriate to base our digital program on a social support-focused theory. We considered all 4 types of social support [48]: *emotional support*, which focuses on sharing life experiences and reassurance; *instrumental support*, which can include practical support such as material or services that users can access directly; *informational support*, which involves evidence-based information or advice; and *appraisal support*, which includes concepts such as constructive feedback and affirmation. Lakey and Cohen's [49] Social Support Theory was adopted to illustrate the relationship between *digital* social support and individuals' appraisals and coping skills, which can, in turn, mediate distress levels and improve overall well-being and T2D self-care.

Table 2. Design of practical applications.

Mechanisms of action (accompanied by BCTs[a])	Practical applications	Text description
Knowledge (information about health consequences, biofeedback, feedback on behavior)	Educational material; Interactive quiz; Animation videos; Decision tracker	Provide information about T2D[b], complications, medication, and healthy lifestyle, including weight, diet, and physical activity; track and receive feedback on health.
Behavioral regulation (self-monitoring of behavior)	Sync steps; Leader board; Track daily targets	App users are encouraged to record dietary habits, weight, HbA_{1c}[c], well-being, blood glucose, cholesterol, and smoking habits, to sync and monitor steps, and complete an action plan for desired behavior.
Skills (coping strategies and habit formation)	Educational material; Interactive quiz; Chat forum	Provide information about a healthy lifestyle; encourage app users to record their tracks; opportunity to share strategies and experiences with other users.
Social and professional role (no BCTs linked to this domain)	Expert videos; Ask the expert	Highly respected experts provide (in the form of videos or emails) and present information and key messages about T2D, well-being, diet, physical activity, and other relevant information.
Beliefs about capabilities (verbal persuasion to boost self-efficacy)	Ask the expert; Chat forum; Patient testimonials	Verbal reassurance and encouragement among users that may boost their confidence and encourage them to continue with their action plan and behavior change
Motivations and goals (goal setting, review of goal, action planning, problem- solving)	Action Plan; Daily targets; Leader board; Personal achievements; Challenges	App users have access to functions and are encouraged to take part in challenges and set their daily targets. They are also encouraged to complete an action plan for target behaviors.
Emotions (reduce negative emotions, social support [emotional])	Chat forum; Whiteboard animation; Educational material; Decision tracker; Quiz	Provide information in the form of whiteboard animations, quizzes, and educational material around burnout, distress, depression, and well-being; use of a chat forum to provide an additional element of peer support.
Social influences (social comparison, social support, encouragement)	Patient testimonials; Chat forum; Leader board; Personal achievements; Challenges	App users have access to videos of peers discussing and sharing experiences of their journey. Chat forums are available for users who wish to share questions and thoughts—this platform acts as a mode for users to encourage one another.

[a]BCT: behavior change technique.

[b]T2D: type 2 diabetes.

[c]HbA_{1c}: glycated hemoglobin.

Step 4: Program Development

Develop Program Content

On the basis of key learnings from our face-to-face program (DESMOND) and our digital programs (BabySteps and myWellbeing) in combination with our theory-based approach, we had a clear idea of the program themes, material list, and aim of MyDESMOND. With the involvement of our multidisciplinary team, we designed scripts and material in line with the Corbin and Strauss self-management framework, incorporating content around medical, role, and emotional management. Once the content was adapted and developed, we liaised with our Web-designer and information technology team to map the material (ie, text, videos, and quiz) onto an appropriate digital platform that was easy to access from any electronic device (ie, desktop computer, smartphone, and tablet). MyDESMOND was developed to meet the objectives detailed in step 3. The following topics are included in MyDESMOND: *What is T2D, Medication, Complications of diabetes, Food choices, Physical Activity, Sedentary behavior, Emotions, Diabetes-related distress, Setting goals, and Relapse*.

Refine Content and Format

In addition to our previous work mentioned in the above sections, it was essential that we also involved patient and public feedback throughout the development of MyDESMOND. Furthermore, PPI work with early users helped inform and refine the content and format of our digital program. PPI work took place based on an iterative process, whereby 9 PPI members were provided logging details to *have a go* at MyDESMOND at various points of the development phase and provide feedback. Examples of PPI feedback and suggestions, which refined MyDESMOND included the following: (1) to provide the chat forum function as optional to the user (and not mandatory), as not all users would seek Web-based interaction with peers; (2) to include a separate section on well-being and discuss about diabetes-related distress and negative emotions in a nonjudgmental way—this suggestion from the PPI members reiterated issues around stigma and blame; (3) to provide weekly additional sessions on well-being, physical activity, and food choices—this change resulted in the development of additional mini videos to ensure that the users received regular information on important topics; and (4) to introduce mini videos of testimonials of MyDESMOND users sharing personal

experience. Additional feedback received from our PPI members included (1) the navigation system of MyDESMOND was easy to use, (2) to make our logo clear to reassure that we are a reliable and trustworthy source; (3) digital ongoing support is much needed poststructured education programs; therefore, all content must be available at all times to the users.

MyDESMOND Platform

The interface and features were designed as such to improve diabetes self-management practices. The focus was directed toward daily step count, challenges, notifications, and the release of weekly education bite-sized sessions. These short sessions consisted of animations, gamification, quizzes, and games designed to be no more than 5 min in duration. Features include interactive learning, monitor health, activity tracking, ask the expert, community (share ideas, experiences, and questions with others in a chat forum), goal setting, challenges (complete challenges with other users on global leader boards to increase steps and focus on personal achievements), and invite a buddy (invite family and friends to join the program). Figures 4-6 present examples of features and description of MyDESMOND content.

Figure 4. Users can learn about T2D and self-management activities and refresh knowledge with 8-weekly additional sessions (educational material, interactive quiz, animations) on MyDESMOND.

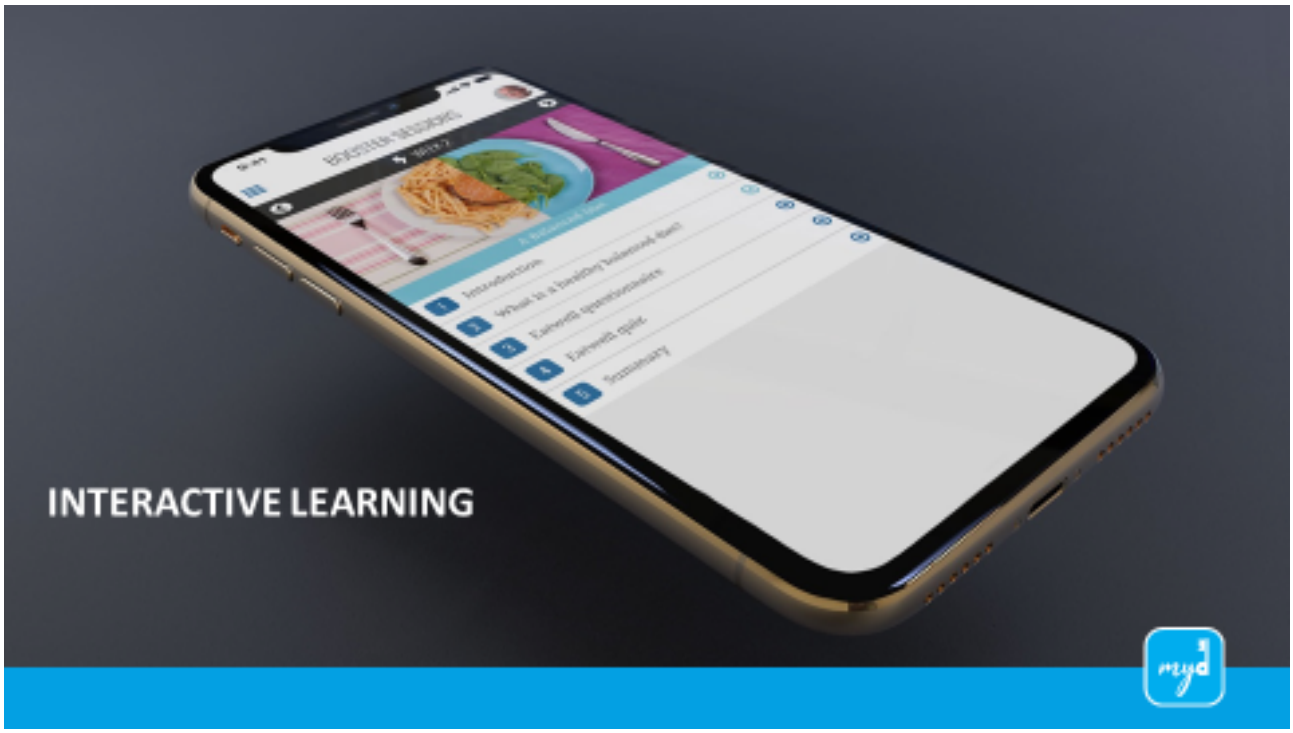
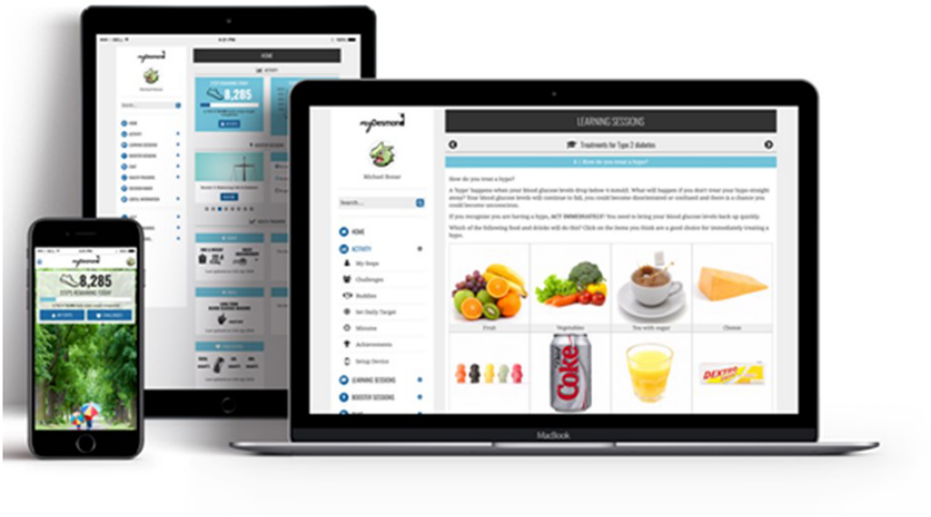


Figure 5. MyDESMOND can track weight, blood pressure, HbA_{1c}, diet, and cholesterol.

Figure 6. MyDESMOND can track activity levels and steps.

MyDESMOND is accessible on any digital device (tablet, desktop computer, and smartphone). The platform was built using progressive Web applications (PWAs), a type of mobile app delivered through the Web. It operates similar to a native app, but its responsive design allows it to work across all modern browsers and devices (smartphone, tablet, and desktop computer; see Figure 7).

Figure 7. MyDESMOND operates like a native app.



The system was developed using the Zend framework [50] and a MySQL database [51]. The app includes an application programming interface built with Zend that is used to manage data between the AngularJS PWA and the database. All identifiable user data, comments, and *Ask the expert* messages are encrypted and stored securely on the database. All data are stored on servers in the United Kingdom and are fully compliant with the latest industry standards for security and General Data Protection Regulations.

Both functional and nonfunctional testing was carried out throughout each step of the development phase by our research team, project-development team, Web-designers, PPI group, and external subject matter experts (SMEs). External SMEs carried out a number of security tests, including penetration testing (to ensure data remains safe) and a source code review (to ensure it complied with Open Web Application Security Project level 2). The purpose of this security testing was to attempt to identify any potential or actual vulnerability to an attack launched across a computer network that could threaten the confidentiality, availability, and integrity of the information being stored and processed.

Discussion

Summary

In this paper, we described the application of the IM framework for the adaptation and development of a digital self-management program, MyDESMOND. This paper brings a unique perspective to digital development, as it illustrated the process from adapting and translating the evidence and content of a face-to-face education program (DESMOND) to developing a digital version of it.

We described the 4 IM steps that were followed to develop MyDESMOND. In step 1 of the IM process, we established a planning group and a patient and public involvement group and described the context of the intervention and program goals. In step 2, we identified objectives and determinants at early design stages to maintain a focus on the strategies adopted. Step 3 focused on generating the program components underpinned by appropriate psychological theories and models. This led to step 4, which described the adaptation and development of the program content and the iterative process of refining the content and format of MyDESMOND.

Strengths and Limitations

This paper has described in detail the three key perspectives of IM that informed the development of MyDESMOND: the use of evidence, the participation and engagement of stakeholders, and the use of theories and BCTs. Overall, our development work was based on a well-established and comprehensive theoretical framework, which has been used extensively in the development of behavior change programs.

The development process was heavily guided by stakeholder and PPI input, to ensure that we received regular feedback throughout the development phase and to ensure that the program was designed appropriately and was relevant to the target population, culturally and contextually. This approach helped refine and structure the format and content of MyDESMOND to reflect the unmet needs of this population accordingly. It is also important to highlight that although our needs assessment and evidence synthesis were predominantly based on findings from systematic reviews, our evidence was also based on rigorous research studies evaluating face-to-face and digital programs.

Some of the main limitations of IM, which we experienced during this development process, are the complexity and time-consuming aspects of this framework. Following a robust process such as IM, which included conducting systematic

reviews, synthesizing evidence from existing studies, obtaining regular feedback from stakeholders, and identifying key theories and BCTs, required substantial time and resources from our team. Nevertheless, we found IM to be an extremely useful resource, which helped develop the best possible digital self-management program for our people with T2D.

Lessons Learned

DHIs in diabetes self-management and other chronic conditions are increasing. It is important that future research recognizes the complexity of DHIs and continues to adopt a robust development process informed by evidence and theory. During the development of MyDESMOND, we identified key learning points that might be helpful to other researchers developing DHIs for chronic conditions.

Multidisciplinary Team

The expertise of our research, project development, and clinical team members was critical to the successful development of the content, including diabetes education, diabetes clinical expertise, and behavior change. Together with experts in Web-design and digital health, the infrastructure of the program was shaped in such a way that it is now usable on NHS systems and maintains the ethical administration of sensitive information. The collaboration between research teams and Web developers was essential so that the focus of development is not only on the content of the program but also on the user interface and functionality of MyDESMOND.

Interface Design

We followed a target population-centered approach, based on the taxonomy of approaches to developing DHI [52]. This approach suggests that any intervention must be informed by the actions and views of the people who are end users. Ongoing involvement of the target group and end users is fundamental for successful intervention design. We ensured that people living with T2D and stakeholders were regularly involved throughout the design of MyDESMOND. This co-design process meant that end users were part of an iterative refinement process to certify that the language and content were easy to understand, and to ensure the navigation system was simple to use.

Theoretical Framework

The role of theories in DHI is crucial, but yet not all programs are guided by theories [53]. It is integral that key theories and models are made explicit in digital programs to enhance understanding of the intended intervention, guide the evaluation process, and inform the implementation of the program in the *real world* [54]. We ensured that MyDESMOND was underpinned by a strong theoretical framework that embedded behavioral theories (HAPA model, Social Support theory, and COM-B model) and the BCT taxonomy to predict and understand behavior change.

Program Functions

Step 1 of the IM process, which included the mixed methods needs assessment, allowed us to collate findings and insight not only from our own work but also from the current literature. This helped identify a number of key functions that were eventually incorporated in MyDESMOND, including information, personal stories, interactions with other app users, and *buddies*. Literature in chronic self-management has an established series of evidence to confirm that providing a secure platform for communication (ie, in the form of chat forums), allows the end user to share personal experiences, increase their knowledge, and improve self-care [55-57]. In addition, literature around gamification suggests that virtual support from friends and family is associated with increased physical activity and that game design and social incentives (*buddies*) can enhance collaboration, accountability, and peer support [58].

Future Directions

The evaluation and implementation of MyDESMOND is currently ongoing in the United Kingdom and Australia, and results on its effectiveness will follow [59]. A follow-up paper on evaluation and implementation (step 5 and step 6 of the IM framework) is also currently in working progress. Despite the focus of the program aiming at adults with T2D, the IM framework could help adapt the current digital platform to new populations and health conditions and provide a taxonomy of behavior change that would inform the content and functionality of new digital programs.

Conclusion

IM framework was an effective approach to bring together research evidence and theoretical concepts to guide the adaptation and development of MyDESMOND. Our evidence- and theory-based digital program provides ongoing support and guidance to people with T2D. We hope that our paper can be used as an example for other researchers and HCPs who are developing similar digital self-management programs across chronic diseases.

Acknowledgments

The authors would like to thank the Web developers and the PPI group for their invaluable input in the development of MyDESMOND. The authors would also like to thank the multidisciplinary team and the individuals who contributed to the testimonial and expert videos. The authors also acknowledge support from the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care East Midlands, the NIHR Leicester Biomedical Research Centre, which is a partnership between University Hospitals of Leicester NHS Trust and the University of Leicester. The program was funded by the Leicester Diabetes Centre.

Authors' Contributions

MD and KK led on the idea of MyDESMOND; CB led on the development of MyDESMOND; AN led on the content development; MH led on the underpinning theories and behavior change; SS led on the implementation of MyDESMOND; and BS led on the content adaptation of MyDESMOND. BS, SS, MH, KK, and MD contributed toward the development of the content of MyDESMOND. MH led the development of the manuscript. SS, BS, AN, CB, KK, and MD revised the content and structure of the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

MD is the principal investigator on the DESMOND programme (Diabetes education and self-management for people with newly diagnosed Type 2 diabetes). KK, SS, CB and AN are employed by the University Hospitals of Leicester NHS Trust, which receives not-for-profit income for DESMOND. All authors are actively engaged in research and have previously received grants for DESMOND from NIHR, Medical Research Council, and Diabetes UK to develop and test Diabetes Self-Management Education and Support programmes such as DESMOND. The University Hospitals of Leicester (UHL) NHS Trust (for which KK, SS, CB and AN have a contract) receives licensing fees to support implementation of the DESMOND programme in Clinical Commissioning Groups in the UK, Ireland and Australia.

References

1. Chen L, Magliano DJ, Zimmet PZ. The worldwide epidemiology of type 2 diabetes mellitus--present and future perspectives. Nat Rev Endocrinol 2011 Nov 8;8(4):228-236. [doi: 10.1038/nrendo.2011.183] [Medline: 22064493]
2. Powers MA, Bardsley J, Cypress M, Duker P, Funnell MM, Fischl AH, et al. Diabetes self-management education and support in type 2 diabetes: a joint position statement of The American Diabetes Association, The American Association of Diabetes Educators, and The Academy of Nutrition and Dietetics. Diabetes Care 2015 Jul;38(7):1372-1382. [doi: 10.2337/dc15-0730] [Medline: 26048904]
3. The National Institute for Health and Care Excellence. 2011 Mar 30. Diabetes in Adults: Quality Standard URL: https://www.nice.org.uk/guidance/qs6/resources/diabetes-in-adults-pdf-58299425989 [accessed 2019-08-19]
4. Davies MJ. ISRCTN Registry. 2011. DESMOND (Diabetes Education and Self Management for Ongoing and Newly Diagnosed) Ongoing Study URL: http://www.isrctn.com/ISRCTN71467886 [accessed 2020-02-28]
5. Gillett M, Dallosso HM, Dixon S, Brennan A, Carey ME, Campbell MJ, et al. Delivering the diabetes education and self management for ongoing and newly diagnosed (DESMOND) programme for people with newly diagnosed type 2 diabetes: cost effectiveness analysis. Br Med J 2010 Aug 20;341:c4093 [FREE Full text] [doi: 10.1136/bmj.c4093] [Medline: 20729270]
6. Deakin TA, Cade JE, Williams R, Greenwood DC. Structured patient education: the diabetes X-PERT Programme makes a difference. Diabet Med 2006 Sep;23(9):944-954. [doi: 10.1111/j.1464-5491.2006.01906.x] [Medline: 16922700]
7. Trento M, Gamba S, Gentile L, Grassi G, Miselli V, Morone G, ROMEO Investigators. Rethink Organization to iMprove Education and Outcomes (ROMEO): a multicenter randomized trial of lifestyle intervention by group care to manage type 2 diabetes. Diabetes Care 2010 Apr;33(4):745-747 [FREE Full text] [doi: 10.2337/dc09-2024] [Medline: 20103547]
8. Ko SH, Song K, Kim S, Lee J, Kim J, Shin J, et al. Long-term effects of a structured intensive diabetes education programme (SIDEP) in patients with Type 2 diabetes mellitus--a 4-year follow-up study. Diabet Med 2007 Jan;24(1):55-62. [doi: 10.1111/j.1464-5491.2007.02013.x] [Medline: 17227325]
9. Chatterjee S, Davies MJ, Heller S, Speight J, Snoek FJ, Khunti K. Diabetes structured self-management education programmes: a narrative review and current innovations. Lancet Diabetes Endocrinol 2018 Feb;6(2):130-142. [doi: 10.1016/S2213-8587(17)30239-5] [Medline: 28970034]
10. National collaborating centre for chronic conditions (UK). Type 2 Diabetes: National Clinical Guideline for Management in Primary and Secondary Care (Update). In: NICE Clinical Guidelines, No. 66. London: NICE; 2008.
11. Winkley K, Evwierhoma C, Amiel SA, Lempp HK, Ismail K, Forbes A. Patient explanations for non-attendance at structured diabetes education sessions for newly diagnosed Type 2 diabetes: a qualitative study. Diabet Med 2015 Jan;32(1):120-128. [doi: 10.1111/dme.12556] [Medline: 25081181]
12. Hagger V, Trawley S, Hendrieckx C, Browne JL, Cameron F, Pouwer F, et al. Diabetes MILES Youth-Australia: methods and sample characteristics of a national survey of the psychological aspects of living with type 1 diabetes in Australian youth and their parents. BMC Psychol 2016 Aug 12;4(1):42 [FREE Full text] [doi: 10.1186/s40359-016-0149-9] [Medline: 27519408]
13. NHS Digital. Diabetes UK. 2019. National Diabetes Audit URL: https://digital.nhs.uk/data-and-information/publications/statistical/national-diabetes-audit/national-diabetes-audit-report-1-findings-and-recommendations-2016-17 [accessed 2019-03-28]
14. Ross J, Stevenson FA, Dack C, Pal K, May CR, Michie S, et al. Health care professionals' views towards self-management and self-management education for people with type 2 diabetes. BMJ Open 2019 Jul 16;9(7):e029961 [FREE Full text] [doi: 10.1136/bmjopen-2019-029961] [Medline: 31315874]

15. NHS Digital. 2017. National Diabetes Audit - 2015-2016: Report 1, Care Processes and Treatment Targets URL: https://digital.nhs.uk/data-and-information/publications/statistical/national-diabetes-audit/national-diabetes-audit-2015-2016-report-1-care-processes-and-treatment-targets [accessed 2019-08-20]
16. NHS Long Term Plan. 2019 Jan. The NHS Long Term Plan URL: https://www.longtermplan.nhs.uk/wp-content/uploads/2019/08/nhs-long-term-plan-version-1.2.pdf [accessed 2019-09-01]
17. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
18. Pal K, Eastwood SV, Michie S, Farmer A, Barnard ML, Peacock R, et al. Computer-based interventions to improve self-management in adults with type 2 diabetes: a systematic review and meta-analysis. Diabetes Care 2014 Jun;37(6):1759-1766. [doi: 10.2337/dc13-1386] [Medline: 24855158]
19. Beatty L, Lambert S. A systematic review of internet-based self-help therapeutic interventions to improve distress and disease-control among adults with chronic health conditions. Clin Psychol Rev 2013 Jun;33(4):609-622. [doi: 10.1016/j.cpr.2013.03.004] [Medline: 23603521]
20. Murray E, Sweeting M, Dack C, Pal K, Modrow K, Hudda M, et al. Web-based self-management support for people with type 2 diabetes (HeLP-Diabetes): randomised controlled trial in English primary care. BMJ Open 2017 Sep 27;7(9):e016009 [FREE Full text] [doi: 10.1136/bmjopen-2017-016009] [Medline: 28954789]
21. Murray E. Web-based interventions for behavior change and self-management: potential, pitfalls, and progress. Med 2 0 2012;1(2):e3 [FREE Full text] [doi: 10.2196/med20.1741] [Medline: 25075231]
22. Hedman E, Andersson E, Ljótsson B, Axelsson E, Lekander M. Cost effectiveness of internet-based cognitive behaviour therapy and behavioural stress management for severe health anxiety. BMJ Open 2016 Apr 25;6(4):e009327 [FREE Full text] [doi: 10.1136/bmjopen-2015-009327] [Medline: 27113231]
23. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. Br Med J 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
24. Bartholomew EL, Parcel GS, Kok G, Gottlieb NH, FernÃ¡ndez ME. Planning Health Promotion Programs: An Intervention Mapping Approach. Fourth Edition. San Francisco, California: Jossey-Bass; 2016.
25. Michie S. Designing and implementing behaviour change interventions to improve population health. J Health Serv Res Policy 2008 Oct;13(Suppl 3):64-69. [doi: 10.1258/jhsrp.2008.008014] [Medline: 18806194]
26. Research Design Service East Midlands. 2017. Public Involvement URL: http://www.rds-eastmidlands.nihr.ac.uk/public-involvement [accessed 2016-08-12]
27. Lorig KR, Holman H. Self-management education: history, definition, outcomes, and mechanisms. Ann Behav Med 2003 Aug;26(1):1-7. [doi: 10.1207/S15324796ABM2601_01] [Medline: 12867348]
28. Strom JL, Egede LE. The impact of social support on outcomes in adult patients with type 2 diabetes: a systematic review. Curr Diab Rep 2012 Dec;12(6):769-781 [FREE Full text] [doi: 10.1007/s11892-012-0317-0] [Medline: 22949135]
29. King DK, Glasgow RE, Toobert DJ, Strycker LA, Estabrooks PA, Osuna D, et al. Self-efficacy, problem solving, and social-environmental support are associated with diabetes self-management behaviors. Diabetes Care 2010 Apr;33(4):751-753 [FREE Full text] [doi: 10.2337/dc09-1746] [Medline: 20150299]
30. Coventry PA, Fisher L, Kenning C, Bee P, Bower P. Capacity, responsibility, and motivation: a critical qualitative evaluation of patient and practitioner views about barriers to self-management in people with multimorbidity. BMC Health Serv Res 2014 Oct 31;14:536 [FREE Full text] [doi: 10.1186/s12913-014-0536-y] [Medline: 25367263]
31. Hadjiconstantinou M, Byrne J, Bodicoat DH, Robertson N, Eborall H, Khunti K, et al. Do web-based interventions improve well-being in type 2 diabetes? A systematic review and meta-analysis. J Med Internet Res 2016 Oct 21;18(10):e270 [FREE Full text] [doi: 10.2196/jmir.5991] [Medline: 27769955]
32. Woolley AK, Hadjiconstantinou M, Davies M, Khunti K, Seidu S. Online patient education interventions in type 2 diabetes or cardiovascular disease: a systematic review of systematic reviews. Prim Care Diabetes 2019 Feb;13(1):16-27. [doi: 10.1016/j.pcd.2018.07.011] [Medline: 30131299]
33. Khunti K, Gray LJ, Skinner T, Carey ME, Realf K, Dallosso H, et al. Effectiveness of a diabetes education and self management programme (DESMOND) for people with newly diagnosed type 2 diabetes mellitus: three year follow-up of a cluster randomised controlled trial in primary care. Br Med J 2012 Apr 26;344:e2333 [FREE Full text] [doi: 10.1136/bmj.e2333] [Medline: 22539172]
34. Leventhal H, Phillips LA, Burns E. The Common-Sense Model of Self-Regulation (CSM): a dynamic framework for understanding illness self-management. J Behav Med 2016 Dec;39(6):935-946. [doi: 10.1007/s10865-016-9782-2] [Medline: 27515801]
35. Durning SJ, Dong T, Artino AR, van der Vleuten C, Holmboe E, Schuwirth L. Dual processing theory and experts' reasoning: exploring thinking on national multiple-choice questions. Perspect Med Educ 2015 Aug;4(4):168-175 [FREE Full text] [doi: 10.1007/s40037-015-0196-6] [Medline: 26243535]
36. Zimmerman BJ, Schunk DH. The scholar and his contributions to educational psychology. In: Educational Psychology: A Century of Contributions. New York: Routledge; 2003:431-457.

37. Anderson RM, Funnell MM, Butler PM, Arnold MS, Fitzgerald JT, Feste CC. Patient empowerment. Results of a randomized controlled trial. Diabetes Care 1995 Jul;18(7):943-949. [doi: 10.2337/diacare.18.7.943] [Medline: 7555554]
38. Sukumar N, Dallosso H, Saravanan P, Yates T, Telling C, Shorthose K, et al. Baby Steps - a structured group education programme with accompanying mobile web application designed to promote physical activity in women with a history of gestational diabetes: study protocol for a randomised controlled trial. Trials 2018 Dec 12;19(1):682 [FREE Full text] [doi: 10.1186/s13063-018-3067-8] [Medline: 30541621]
39. Hadjiconstantinou M. British Library EThOS. UK; 2018. The Development of an Online Emotional Support Programme to Improve Diabetes-Related Distress and Well-Being in Type 2 Diabetes URL: https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.745902 [accessed 2019-09-01]
40. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
41. Pound P, Campbell R. Exploring the feasibility of theory synthesis: a worked example in the field of health related risk-taking. Soc Sci Med 2015 Jan;124:57-65 [FREE Full text] [doi: 10.1016/j.socscimed.2014.11.029] [Medline: 25461862]
42. Schulman-Green D, Jaser S, Martin F, Alonzo A, Grey M, McCorkle R, et al. Processes of self-management in chronic illness. J Nurs Scholarsh 2012 Jun;44(2):136-144 [FREE Full text] [doi: 10.1111/j.1547-5069.2012.01444.x] [Medline: 22551013]
43. Michie S, Atkins L, West R. The Behaviour Change Wheel: A Guide To Designing Interventions. UK: Silverback; 2014.
44. Atkins L, Francis J, Islam R, O'Connor D, Patey A, Ivers N, et al. A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. Implement Sci 2017 Jun 21;12(1):77 [FREE Full text] [doi: 10.1186/s13012-017-0605-9] [Medline: 28637486]
45. French SD, Green SE, O'Connor DA, McKenzie JE, Francis JJ, Michie S, et al. Developing theory-informed behaviour change interventions to implement evidence into practice: a systematic approach using the Theoretical Domains Framework. Implement Sci 2012 Apr 24;7:38 [FREE Full text] [doi: 10.1186/1748-5908-7-38] [Medline: 22531013]
46. Michie S, Johnston M, Francis J, Hardeman W, Eccles M. From theory to intervention: mapping theoretically derived behavioural determinants to behaviour change techniques. Appl Psychol 2008;57(4):660-680. [doi: 10.1111/j.1464-0597.2008.00341.x]
47. Webb TL, Sheeran P. Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. Psychol Bull 2006 Mar;132(2):249-268. [doi: 10.1037/0033-2909.132.2.249] [Medline: 16536643]
48. House JS. Social support and social structure. Sociol Forum 1987;2(1):135-146. [doi: 10.1007/bf01107897]
49. Lakey B, Cohen S. Social support theory and measurement. In: Cohen S, Underwood LG, Gottlieb BH, editors. Social Support Measurement and Intervention: A Guide for Health and Social Scientists. Oxford, UK: Oxford University Press; 2000:29-52.
50. Pope K. Zend Framework 1.8 Web Application Development. UK: Packt Publishing; 2009.
51. Greenspan J, Wall D, Bulger B. MySQL / PHP Database Applications. USA: John Wiley & Sons; 2001.
52. O'Cathain A, Croot L, Sworn K, Duncan E, Rousseau N, Turner K, et al. Taxonomy of approaches to developing interventions to improve health: a systematic methods overview. Pilot Feasibility Stud 2019;5:41 [FREE Full text] [doi: 10.1186/s40814-019-0425-6] [Medline: 30923626]
53. Wang Y, Fadhil A, Lange J, Reiterer H. Integrating taxonomies into theory-based digital health interventions for behavior change: a holistic framework. JMIR Res Protoc 2019 Jan 15;8(1):e8055 [FREE Full text] [doi: 10.2196/resprot.8055] [Medline: 30664477]
54. Murray E, Treweek S, Pope C, MacFarlane A, Ballini L, Dowrick C, et al. Normalisation process theory: a framework for developing, evaluating and implementing complex interventions. BMC Med 2010 Oct 20;8:63 [FREE Full text] [doi: 10.1186/1741-7015-8-63] [Medline: 20961442]
55. Powell JA, Darvell M, Gray JA. The doctor, the patient and the world-wide web: how the internet is changing healthcare. J R Soc Med 2003 Feb;96(2):74-76 [FREE Full text] [doi: 10.1258/jrsm.96.2.74] [Medline: 12562977]
56. Nambisan P. Information seeking and social support in online health communities: impact on patients' perceived empathy. J Am Med Inform Assoc 2011 May 1;18(3):298-304 [FREE Full text] [doi: 10.1136/amiajnl-2010-000058] [Medline: 21486888]
57. van der Eijk M, Faber MJ, Aarts JW, Kremer JA, Munneke M, Bloem BR. Using online health communities to deliver patient-centered care to people with chronic conditions. J Med Internet Res 2013 Jun 25;15(6):e115 [FREE Full text] [doi: 10.2196/jmir.2476] [Medline: 23803284]
58. Patel MS, Benjamin EJ, Volpp KG, Fox CS, Small DS, Massaro JM, et al. Effect of a game-based intervention designed to enhance social incentives to increase physical activity among families: The BE FIT Randomized Clinical Trial. JAMA Intern Med 2017 Nov 1;177(11):1586-1593 [FREE Full text] [doi: 10.1001/jamainternmed.2017.3458] [Medline: 28973115]
59. MyDesmond | Supporting you to self-manage your diabetes. URL: https://www.mydesmond.com/ [accessed 2019-09-01]

Abbreviations

BCT: behavior change techniques
DHI: digital health intervention
HAPA: Health Action Process Approach
HCP: health care professionals
IM: intervention mapping
NHS: National Health Service
NIHR: National Institute for Health Research
PPI: patient and public involvement
PWA: progressive Web application
SME: subject matter experts
T2D: type 2 diabetes

Edited by G Eysenbach; submitted 05.12.19; peer-reviewed by A Lake, J Parsons; comments to author 19.01.20; revised version received 31.01.20; accepted 07.02.20; published 11.05.20.

Please cite as:
Hadjiconstantinou M, Schreder S, Brough C, Northern A, Stribling B, Khunti K, Davies MJ
Using Intervention Mapping to Develop a Digital Self-Management Program for People With Type 2 Diabetes: Tutorial on MyDESMOND
J Med Internet Res 2020;22(5):e17316
URL: https://www.jmir.org/2020/5/e17316
doi:10.2196/17316
PMID:32391797

Original Paper

A Facilitated Web-Based Self-Management Tool for People With Type 1 Diabetes Using an Insulin Pump: Intervention Development Using the Behavior Change Wheel and Theoretical Domains Framework

Claire Reidy[1,2], PhD; Claire Foster[3], PhD; Anne Rogers[1], PhD

[1]National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care, School of Health Sciences, Faculty of Environmental & Life Sciences, University of Southampton, Southampton, United Kingdom

[2]School of Primary Care, Population Health & Medical Education, Faculty of Medicine, University of Southampton, Southampton, United Kingdom

[3]Macmillan Survivorship Research Group, School of Health Sciences, Faculty of Environmental & Life Sciences, University of Southampton, Southampton, United Kingdom

Corresponding Author:
Claire Reidy, PhD
National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care
School of Health Sciences, Faculty of Environmental & Life Sciences
University of Southampton
University Road
Southampton, SO17 1BJ
United Kingdom
Phone: 44 23 8059 7628
Email: c.reidy@soton.ac.uk

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2020/7/e21381/

Abstract

Background: Type 1 diabetes (T1D) requires intensive self-management (SM). An insulin pump is designed to better support personal T1D management, but at the same time, it exacerbates the complexity and requirements of SM. Research shows that people with diabetes are likely to benefit from navigating and connecting to local means of social support and resources through web-based interventions that offer flexible, innovative, and accessible SM. However, questions remain as to which behavior change mechanisms within such resources benefit patients most and how to foster engagement with and endorsement of SM interventions.

Objective: The aim of this study was to evaluate the perspectives and experiences of people with T1D using an insulin pump and specialist health care professionals (HCPs) and determine what behavior change characteristics and strategies are required to inform the optimization of an existing web-based social network (SN) intervention to support SM.

Methods: Focus groups with insulin pump users (n=19) and specialist HCPs (n=20) in 6 National Health Service (NHS) trusts across the south of England examined the barriers and enablers to incorporating and self-managing an insulin pump. An analysis was undertaken using the Behavior Change Wheel and Theoretical Domains Framework, followed by a taxonomy of behavior change techniques (BCTs) to identify the contents of and strategies for the implementation of a complex health intervention.

Results: A total of 4 themes represent the SM perspectives and experiences of stakeholders: (1) a desire for access to tailored and appropriate resources and information—the support and information required for successful SM are situational and contextual, and these vary according to time and life circumstances, and therefore, these need to be tailored and appropriate; (2) specific social support preferences—taking away isolation as well as providing shared learnings and practical tips, but limitations included the fear of judgment from others and self-pity from peers; (3) the environmental context, that is, capacity and knowledge of pump clinic HCPs—HCPs acknowledge the patient's need for holistic support but lack confidence in providing it; and (4) professional

responsibility and associated risks and dangers, whereas HCPs are fearful of the consequences of promoting non-NHS SM support, and they question whether SM support fits into their role. BCTs were identified to address these issues.

Conclusions: The use of behavioral theory and a validated implementation framework provided a comprehensive approach for systematically identifying barriers and enablers of self-managing T1D with an insulin pump. A web-based SN intervention appears to offer additional forms of SM support while complementing NHS services. However, for intervention implementation, HCPs' apprehensions about responsibility when signposting to non-NHS SM support would need to be addressed, and opportunistic features would need to be added, through which pump users could actively engage with other people living with T1D.

(J Med Internet Res 2020;22(5):e13980) doi:10.2196/13980

KEYWORDS

web-based intervention; behavior change wheel; type 1 diabetes; social support; continuous subcutaneous insulin infusion; self-management

Introduction

Background

In the United Kingdom, approximately 400,000 people are currently living with type 1 diabetes (T1D), and both the prevalence of T1D and the health care costs of managing T1D are increasing [1,2]. An improvement in blood glucose levels is viewed as a primary goal of self-management (SM) efforts, as it delays the onset and progression of diabetes-related complications (stroke, heart disease, and neuropathy). However, only 30% of the people with T1D are achieving the recommended glycemic targets [3], and attaining these targets is complex. There is recognition of the need for more tailored interventions to enhance the opportunity to improve blood glucose levels [4]. Theoretically founded web-based interventions in particular are seen to offer the opportunity to support flexible, innovative, and accessible SM to address this growing crisis [5].

The Complexities of Treating and Managing Diabetes

Treatment of T1D comprises demanding SM requirements, including insulin therapy (multiple daily injections [MDI] or insulin pumps); self-monitoring of blood glucose; and comprehensive understanding of nutritional, hormonal, and physical impacts on glycemia [6,7]. MDI is the most common insulin therapy method, but interest in and uptake of insulin pumps have risen over the past 20 years, and predictions suggest that this will continue because of the growing global interest and evidence supporting their use [8]. The Diabetes Attitudes, Wishes, and Needs second study (DAWN2) found that the outcomes are better for people with diabetes when they have greater access to diabetes SM education and positive social support [9]. A recent review suggested that interventions to improve these aspects are necessary and require more flexible and personal SM support for those using these devices [10]. The review highlighted how the process of incorporating an insulin pump often changes treatment expectations and experiences and comprises a distinct and potentially difficult process of learning, exploration, and adaptation. People with T1D initiating a new health technology need to self-manage, but they need appropriate options to do so, and web-based interventions have unlocked potential in this regard.

New Approaches to Self-Management of Type 1 Diabetes

Technology can play a key role in bringing diabetes care to the individual [11]. Interest in web-based SM interventions has increased over the past decade [11,12], as web-based elements (or electronic health) offer opportunities to take pressure off the National Health Service (NHS) while supporting flexible and accessible SM [5]. In addition, interventions that take into account the individual's social context in behavior change are relevant in improving health outcomes [13]. It is well recognized that poor psychological well-being can have a significant impact on glycemic control, which consequently increases the risk of diabetes-related complications and leads to increased health care costs and lost productivity [11,14-18]. Increased valued social involvement is linked to greater SM capacity and potentially lower formal health care costs, especially when this involvement is from a diverse set of network members [19,20]. Network members can be health care professionals (HCPs), family, friends, colleagues, community groups, objects (eg, a bicycle), pets, and spiritual groups. Social networks (SNs) and good social support have been shown to promote diabetes SM and assist in physical and mental well-being [21-27]. An SN approach focuses on available and underused collective support from network members as well as on behavior change at a cognitive level [21-27]; therefore, a web-based social support network intervention could provide a currently underutilized avenue for improved psychological well-being and blood glucose levels.

Blakeman et al's [28] randomized controlled trial of an early version of the Generating Engagement in Networks InvolvEment (GENIE) SN intervention demonstrated improved quality of life, engagement in health care, and health outcomes. GENIE is a web-based SN tool that provides SM support by helping participants map their personal community of support and make the best use of existing contacts and add new ones where needed, as well as signposting (and providing a nudge) to personalized resources in their locality [29]. Despite this demonstration of success, little progress has been made in implementing and spreading psychosocial or social support interventions, in general, into clinical practice to improve SM [6,30-33]. There are challenges in the implementation, sustainability, and accessibility of these interventions in local contexts and to relevant stakeholders (patients and HCPs) [34]. The consideration of the mechanisms of success is often missing

[35]. For example, Mulvaney et al's [36] review of the diabetes mobile intervention design found that there was often little consideration for what SM barriers were addressed or the likely motivation for potential users. They suggested tailoring health intervention content and design (such as GENIE) to stakeholder characteristics to improve patient engagement and outcomes.

The Medical Research Council has identified the importance of utilizing theory and incremental stepped approaches when developing behavior change interventions [37]. In this instance, the Behavior Change Wheel (BCW) and the Theoretical Domains Framework (TDF) were selected because of their focus on the context (the physical setting) in which a behavior occurs, the reflective processes that are involved in behavior change, and the provision of a clear and direct strategy to foster change [38-40]. The evidence base for digital SM interventions in long-term conditions may be able to progress more effectively if we not only focus on measured outcomes but also document and examine the dimensions and processes of interventions most important to stakeholders.

This paper provides a comprehensive needs identification of the specific insulin pump SM needs and perspectives of people with T1D and HCPs working in T1D pump clinics. This will identify recommendations to adapt and optimize the preexisting web-based intervention GENIE, both in terms of the content of the existing intervention and the implementation of the intervention with the aim of improving the SM of people with a long-term condition (such as diabetes) while implementing a new health technology (such as an insulin pump) from the point of technology initiation.

Methods

Study Design

This qualitative study consisted of focus groups, which provided the opportunity to explore the range of views and perspectives of the support required and resources used by current pump users, from pump initiation to current point of use. Focus groups were used as a means to facilitate discussion [41], and focus groups are known to stimulate enhanced disclosure and a supportive environment, which incites elaborated accounts and clarification of experiences [42]. Focus groups with HCPs allowed for the exploration of how a web-based SM support tool could fit into NHS practice. The group environment was considered a strength for discussions of implementation and offered an opportunity for individual HCPs to respond to and build on colleague's comments and brainstorm ideas. Focus groups were undertaken until "saturation" (ie, no significant new insights emerged) [43].

Ethics

Ethical approval for this study was granted by the University of Southampton (Reference 26208) and the National Research Ethics Service (Reference 17/NS/0089).

Setting

The study took place between July 2017 and January 2018 in the south of England. The focus groups took place within 6 NHS trusts, which represented varying levels of deprivation and population density across the region.

Population Sample

Pump user participants were purposefully sampled to ensure a range of pump user ages, length of diagnosis, marital status, sex, and employment status to reflect differing perspectives. Clinics were purposively sampled to represent natural variation across different secondary care settings (urban/semirural, varying deprivation levels, and commissioning procedures). Participants in the focus groups held an advisory capacity for the adaption of an SN SM intervention; therefore, variation was prioritized to improve the likelihood of the resulting intervention being fit for purpose, and it was developed appropriately according to the needs of a variety of pump users and within the context of secondary care.

Eligible patient participants were aged 18 to 65 years, had been diagnosed with T1D for more than 1 year, and had an insulin pump for more than 6 months. Participants who had lived with a pump for less time were excluded to focus on the experiences of overcoming, and reflection of, the initial period of adjustment. Diagnosis of diabetes for less than 1 year was also excluded so as to not obscure the experiences of incorporating a new technology with those of a new diagnosis. Participants were invited to take part through social media, posters in local pump clinics, local diabetes charities, and peer support groups.

All HCPs in insulin pump clinics working directly with patients were eligible to participate in the study and were invited to attend focus groups through direct contact with the clinic.

Theory

The BCW [39] is an overarching framework from a synthesis of behavior change interventions providing a clear all-encompassing model of behavior change (see Figure 1). This synthesis integrates theoretical constructs leading to successful behavior change in a variety of health settings. The central cog of the BCW comprises the Capability, Opportunity, Motivation-Behavior (COM-B) components (see Figure 2). This is based on the premise that to initiate behavior change, there is a need to maximize physical or psychological "Capability" to regulate behavior (ie, develop relevant skills), increase or decrease automatic or reflective "Motivation" to engage in desired/undesired behavior, and target the physical or social "Opportunity" to support behavior change. The COM-B offers an understanding of the barriers and enablers of behavior and underscores the potentially modifiable factors for an intervention to target. The BCW links the COM-B model results with intervention functions (see Figure 3). We also utilized the TDF [38,44] (see Figure 2) to provide specific and comprehensive behavioral domains to target in the intervention. The TDF compounds 84 constructs from multiple psychological theories (motivational, action, and organizational theories) and comprises 14 domains of theoretical constructs [44-46]. The TDF provides a useful framework for understanding the barriers and factors influencing specific behaviors [44,47,48]. It provides a detailed analysis of the potentially modifiable factors linked with the BCW (the COM-B components in the central cog of the wheel) to target in an intervention. For example, if lack of

knowledge prevents SM, this would be coded as "psychological capability" in COM-B; thereafter, more specifically, "Knowledge" using the TDF and the intervention function mapping of the BCW might suggest an intervention function of "education." Using the BCW and the TDF in this way has been recommended elsewhere [38,44,47].

A taxonomy of behavior change techniques (BCTs) [49] then enables the specification of techniques describing the active components of the intervention to tailor and optimize an SN intervention. Focus group interview topic guides for both patients and HCPs were developed in consideration of the components of the COM-B model [39] and TDF [44] to ensure participants had the opportunity to explore each element (eg, physical opportunity to self-manage).

Figure 1. Determining the potential mechanisms of action of an intervention using the Behavior Change Wheel.

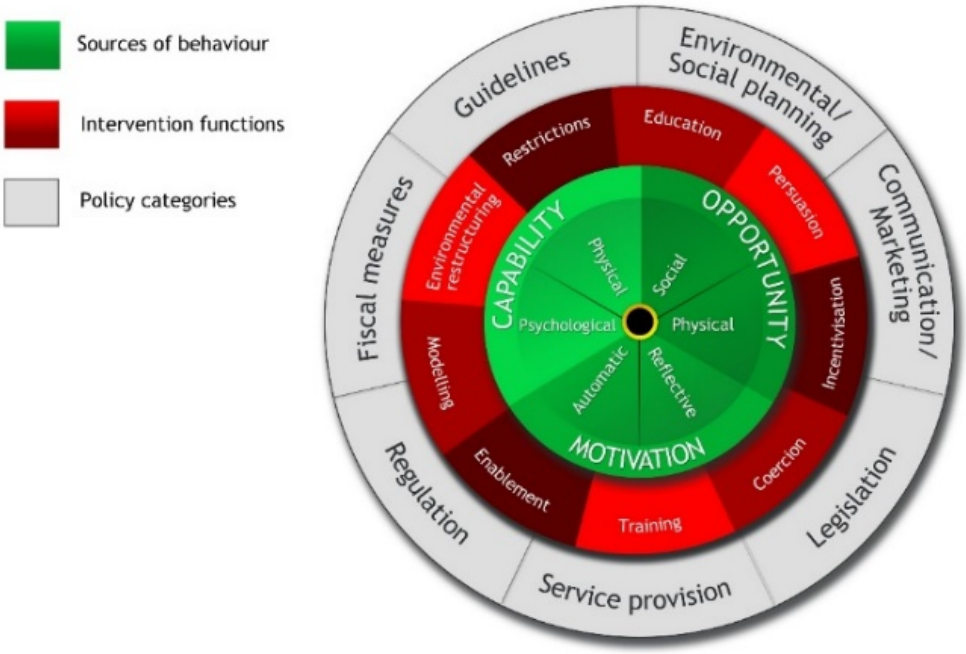


Figure 2. The 14 domains of the Theoretical Domains Framework, structured according to the Capability, Opportunity, Motivation-Behavior model. TDF: Theoretical Domains Framework; Soc: Social influences; Env: Environmental context and resources; Id: Social/professional role and identity; Bel Cap: Beliefs about capabilities; Opt: Optimism; Int: Intentions; Bel Cons: Beliefs about consequences; Reinf: Reinforcement; Em: Emotion; Know: Knowledge; Cog: Cognitive and interpersonal skills; Mem: Memory, attention, and decision processes; Beh Reg: Behavioral regulation; Phy: physical skills.

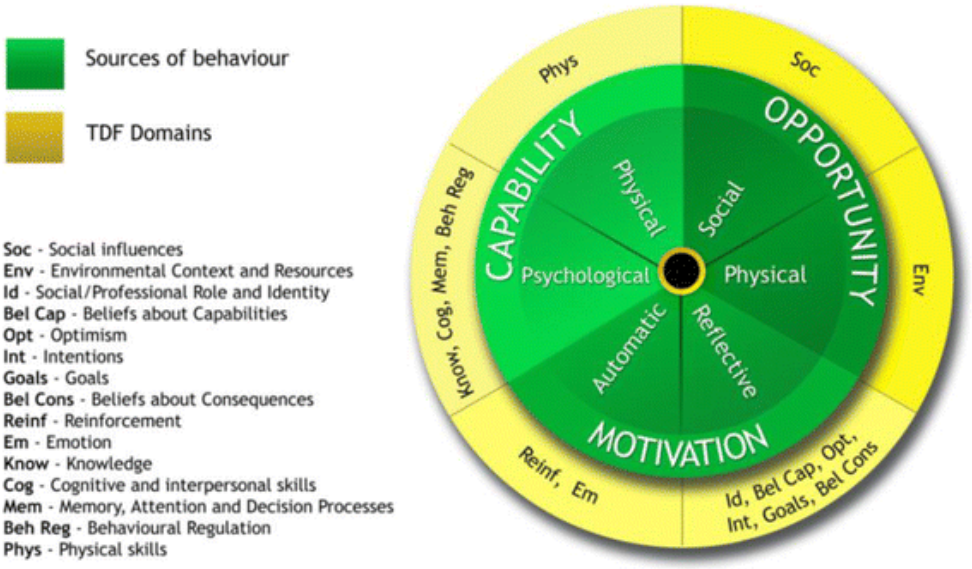
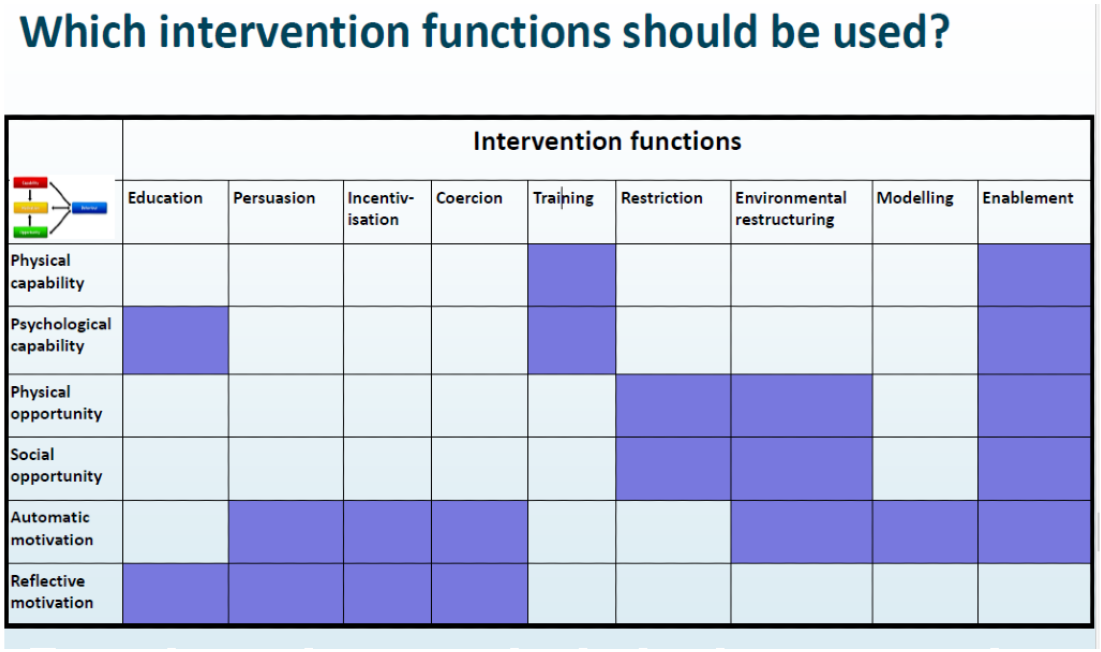


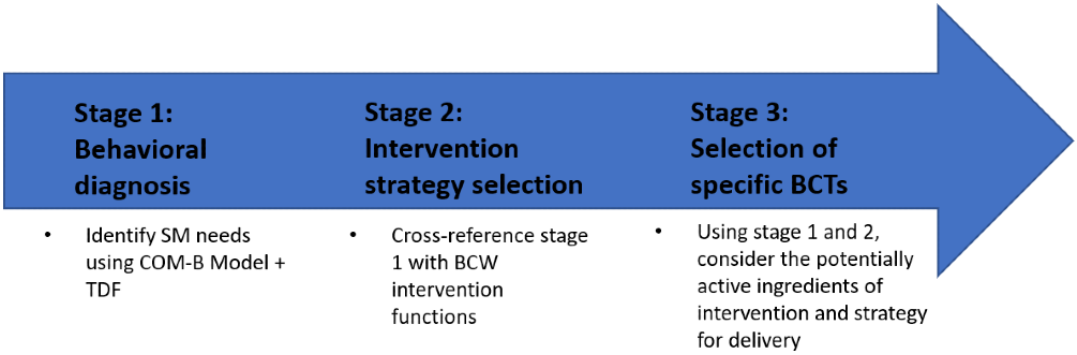
Figure 3. Intervention function mapping matrix.



Behavioral Analysis

Behavioral analysis using the BCW and TDF and complimentary taxonomies of BCTs comprised 3 stages to systematically determine the necessary mechanisms of action for supporting SM and developing a suitable intervention (see Figure 4).

Figure 4. Determining the potential mechanisms of action of an intervention using the Behavior Change Wheel. BCTs: behavior change techniques; SM: self-management; COM-B: Capability, Opportunity, Motivation-Behavior; TDF: Theoretical Domains Framework; BCW: Behavior Change Wheel.



Stage 1: Behavioral Diagnosis

The first stage identifies the behaviors identified in the focus groups in context of using the COM-B model as a framework (stage 1a), broken down into physical or psychological capability, reflective or automatic motivation, and social or physical opportunity and determined as barriers and enablers. The COM-B model was then used alongside the TDF to provide a more comprehensive behavioral analysis by allocating more specific behavioral domains to focus on within the intervention (or delivery of the intervention, eg, a focus on addressing knowledge or skill; stage 1b). All focus group interviews were transcribed verbatim. A deductive approach to analysis was used for the initial analysis, using the theoretical framework provided by the COM-B model and the TDF [50]. The data were further analyzed inductively (by CR) to identify the overarching themes within the COM-B and TDF subcomponents to summarize quotes representing similar underlying ideas (see Figure 3) [51].

Stage 2: Intervention Strategy Selection

The second stage cross-references the behavioral diagnosis (stage 1a and 1b) with the BCW "intervention functions" ("education," "persuasion," "incentivization," "coercion," "training," "restriction," "environmental restructuring," "modeling," and "enablement"; see Figure 1).

The intervention mapping matrix (Figure 3) was employed to establish which intervention functions would be most pertinent in targeting the SM support required.

Stage 3: Selection of Specific Behavior Change Techniques

Stage 3 identified the BCTs that would be required in the facilitated web-based SN intervention—GENIE (components of the intervention such as goal setting, restructuring the social environment, and framing/reframing) [49]—according to the findings of stages 1 and 2. This allowed us to determine the necessary mechanisms of action for SM support intervention

(GENIE) and to which the Acceptability, Practicability, Effectiveness/Cost-effectiveness, Affordability, Safety, Equity (APPEASE) criteria were applied. The APPEASE criteria provide guidelines to consider for the content and delivery of the intervention, based on affordability, practicability, effectiveness and cost-effectiveness, acceptability, side-effects/safety, and equity considerations. In addition, a distinction was made regarding both the potentially active ingredients of the intervention (named "Reflective" processes) and the components and delivery of the intervention (context/setting; named "Strategic" processes).

Results

Results of Focus Groups

A total of 11 focus groups and 1 interview were carried out; 6 focus groups were carried out with insulin pump users (n=19; see Table 1), and 5 focus groups and 1 interview were carried out with diabetes specialist HCPs (n=20). We held 1 focus group per clinic (except one where we also undertook an interview). Conversations lasted from 40 to 72 min (average=56.33 min) with patients and 27 to 44 min (average=37.6 min) with HCPs.

Table 1. Participant demographics.

Characteristics	Values
Age (years), mean (SD); range	38.53 (9.91); 20-53
Sex (female), n (%)	10 (53)
Ethnicity (white British), n (%)	16 (84)
Income[a], n (%)	
Lower than average	8 (42)
Average	6 (32)
Higher than average	5 (26)
Education level (degree level or above), n (%)	12 (63)
Time since diagnosis (years), mean (SD); range	21.95 (12.77); 3-41
Time since pump start (years), mean (SD); range	5.94 (5.98); 0.5-19
Diabetes-related complications[b], n (%)	9 (47)
Been in hospital >3 times[c] for hypoglycemia or diabetic ketoacidosis, n (%)	2 (11)
Health care professionals, n	
Role, n (%)	
Diabetes specialist dietician	5 (25)
Diabetes specialist nurse	7 (35)
Diabetes consultant	7 (35)
Diabetes assistant practitioner	1 (5)
Sex (female), n (%)	15 (75)
Age (years), (%); range	70; 45-54
Ethnicity (white British), n (%)	16 (80)
Time in diabetes clinical practice, mean (SD); range	13.69 years (8.22); 2 months-27 years
Time working with pumps, mean (SD); range	8.74 years (5.98); 2 months-24 years
Time working in current diabetes clinic, mean (SD); range	10.11 years (7.62); 2 months-25 years

[a]Average income in the United Kingdom=£26,500.

[b]Eye damage: background retinopathy/eye damage/treated retinopathy/nerve damage (neuropathy)/other complications.

[c]Over the last 3 years.

Behavioral Analysis Stage 1 Results: Framework Analysis

The findings of the framework analysis providing a matrix of links among the COM-B model and the TDF are presented in Multimedia Appendices 1 and 2.

Stage 1 Results: Thematic Analysis

A total of 4 key themes were identified from the thematic analysis of transcripts; data and quotes are presented to illustrate each theme rather than theoretical subcomponent for conceptual accessibility: (1) desire for access to tailored and appropriate resources and information, (2) specific social support preferences, (3) the environmental context—capacity and

knowledge of pump clinic HCPs, and (4) professional responsibility and associated risks and dangers.

Desire for Access to Tailored and Appropriate Resources and Information

It was acknowledged that at the initiation of pump therapy, the pump can be complicated and difficult to master. Patients reported a desire for holistic support and flexible, convenient access to information and resources as well as access to the latest scientific research, but only at a time suitable for them (as and when). Web-based support was particularly salient because of ease of access. This kind of support, information, and resource was desired in times of heightened difficulty and situational change (eg, pregnancy, health complications, new employment arrangements, and experience of "burnout"). People's time was also limited; therefore, resources had to be used wisely, both in terms of attending clinics and accessing assistance. All the pump user focus groups included substantive discussions about access to tailored and advanced fitness-related information. Performing exercise along with others living with T1D or seeking advice from others about exercise were expected to ease some of the anxieties about experiencing (or preparing for) low (or high) blood glucose levels during exercise:

> *I don't know if any of you have heard of the website Runsweet or Ex-carbs or anything like that?...All of the rest of the Type 1 diabetes management was fine for me, but exercise was my big issue...Anyway, Ex-carbs is a website that helps you to come up with a good way to begin exercising.* [Dan, pump user]

In addition, relevant information was needed, which was specific to T1D and/or insulin pumps, rather than general information for any type of diabetes:

> *It would be nice to have access to a website that gives you information about diets and Type 1 diabetes. I go to [diabetes charity], but it's not up-to-date. It's for Type 2.* [Katherine, pump user]

Access to other holistic pursuits were cited as important, owing to the participants' desire for enjoyable activities for the promotion of positive mental health and/or finding that these activities also required some navigation in terms of the impact on their glycemic control:

> *I've never been really sporty...I also do get a little bit annoyed that every time anyone [in diabetes groups] does talk about any kind of social interactions, other than "meet-ups", it's always revolved around sports. I would love to see, or even run, some more diabetic-friendly groups that are, for example, theatre based. The pressure of being on stage is likely to cause hypos or have a high so you need a group which understands that, you know?* [Stephanie, pump user]

Specific Social Support Preferences

Social support was fundamental to most insulin pump users. Flexible and open contact with the clinic was valued, although this did depend on personal experiences with HCPs, but support from peers was equally valued. Being among other people with T1D, both on the web and offline, provided a wealth of otherwise unseen yet vital information for day-to-day life, such as practical tips and provision of assistance (faulty equipment and where to place the pump on the body). This need varied according to circumstance: T1D-specific support groups, especially if newly diagnosed, were desired, and diabetes-specific fitness groups were valued for the opportunity to determine how best to exercise without glucose levels rising too high or falling too low or how/where to carry the extensive equipment. Meeting peers was associated with taking away some of the isolation of living with a hidden condition:

> *[I would like] social things like groups that you can meet people who are in a similar situation to you...because you can't just walk down the street and ask "are you on a pump?"* [Mark, pump user]

> *But actually I had no idea that diabetes-- I remember thinking this condition was incredibly rare, because I never knew anyone else with it.* [Jenny, pump user]

Access to peer support was cited as important in sharing stories, troubleshooting, sharing illness burden, and speaking to people who understand this "invisible" condition. Some desired web-based support, whereas others desired face-to-face contact, and it was common to desire both. Although face-to-face interactions were important, web-based access allowed people to conveniently "dip in" or "lurk" from a safe distance. In addition, participants expressed wanting to be of assistance themselves, providing support of mutual (reciprocal) benefit. However, apprehensions were raised about accessing people who were in a similar situation. Identities began to be focused on the basis of being pump users:

> *Personally, I find having a one-on-one conversation with someone and asking questions...as wonderful as the nurses are, and the clinic nurses are fantastic, but having someone who uses a pump every single day-It was really positive being in a group setting and having conversations amongst ourselves...You could say "what do you do while you're asleep?" "Do you ever get over having something strapped to you?" Just basic questions.* [Harry, pump user]

> *I guess more links...I had some like issues with it [the pump] sticking on--and no one's ever told me about what kind of tapes that I can use to keep it on or stuff like that, or even nice covers for your pump, just like nice things that are easier to find through that [social-network intervention] rather than having to go through Amazon.* [Lauren, pump user]

There were distinct barriers to speaking to others with T1D, such as a lack of confidence, especially when there was a perceived risk of peer judgment or competition:

> *Because if you are nervous of -- If you don't have the best control or you have been through a bit of a rough patch, or you don't really know-you know-It must be daunting to meet other people so I think you have to be in the right kind of place to want to—* [Jenny, pump user]

The Environmental Context: Capacity and Knowledge of Pump Clinic Health Care Professionals

Many HCPs were positively encouraging of the psychosocial needs of patients and recognized that social and peer support were valuable for patients:

> *Yes, so, it is useful. It's very positive. The good thing I like about it is the opportunity to meet other people, network and do other things outside of diabetes, and for them to feel as normal as possible, but they are normal. You know what I mean?* [Diabetes Specialist Nurse 5, HCP]

HCPs were enthusiastic about supporting their patients to self-manage, especially in terms of patient's need for holistic support and resources, but they lacked confidence in addressing the psychosocial needs of patients themselves:

> *I think it's a question of whether we think we're skilled. I think it's more a part of taking history but it realms into the psychological support, psychology support territory and whether as nurses and dietitians and clinicians, we think we would have the skills to deliver that. I think it's something which if it was something very, do tick box; A, B and C, this is something which we don't do in our routine clinical basis...but a lot of the care is focused towards the more technical and medical and other supportive aspects.* [Consultant 4, HCP]

However, some HCPs expressed a lack of value for psychosocial support or SM support where it was not seen as part of their clinical remit:

> *So, realistically...resources that are available are something that you kind of say--, "oh look I know I've got my little 'talking change' thing" and my "little thing in there for somebody who" and "that's a resource that I can make available", but, I don't say, "Would you like to talk to a psychology person--?" to everybody that comes in…and I suppose that a lot of it is that if it's not broken what's to fix?* [Consultant 6, HCP]

Most clinicians were interested in innovative ways for patients to access other support. They were especially enthusiastic about their patient's needs, with an appreciation of the benefits of engagement with other people with T1D, especially others with a pump, for shared learnings and experiences. Some clinicians considered the potential facilitation of access to social support interventions in structured education sessions, whereas others considered approaches to such support via signposting through their clinic rather than providing access within. However, HCPs were concerned about competing priorities and the consequential lack of time/capacity in the clinic to engage in SM support or offer a facilitated web-based intervention:

> *...I think the CCG fund the pumps but we don't have an awful lot of funding for the team that supports the pump service, so whilst we had small numbers we could incorporate it into our service level agreement but as the pump service has grown we're struggling to offer the support we would like to offer. The feedback we're getting is our pump patients love our service and want more of it but actually we can't really give them anymore because we're not funded to.* [Dietician 4, HCP]

Professional Responsibility and Associated Risks and Dangers

Some HCPs were evidentially concerned about the risk and dangers of signposting or onward referral to a web-based SM support tool, and they held fears that such signposting to a nonclinical environment could have negative consequences in terms of their professional responsibility:

> *Yes, or, accuracy of...Or the potential dangers of peer-to-peer advice regarding immediate clinical matters. I think that's my opinion at the moment. Sharing it in a controlled way with the, you know, organizations that are available to have them. In terms of peer-to-peer advice, what if someone gives them the wrong advice? Maliciously, for instance.* [Consultant 7, HCP]

Some HCPs also felt that this could be "creating problems for problems sake" by offering SM support services within a clinical setting:

> *My first thought about this, is it bringing up things that we actually don't need to bring up, I would think that. I know we do want to make sure that everybody is well supported and has access to that support. At the same time, if somebody's absolutely fine...We don't want to be making them feel that there is something wrong when there isn't...What you don't want to be doing is creating problems. For problems sake.* [Dietician 3, HCP]

However, pump users referred to unhelpful experiences of HCPs blocking access to information, resources, or medical equipment. Patients demonstrated an understanding of risk, but they also demonstrated the need to make decisions themselves:

> *Going back to that idea of online groups, I understand that you would want to have a warning to say, "this is not NHS, this is not moderated. This is just a group that is publicly available and we're not recommending or making any sort of judgment". I'm fine with the warning but ideally would want to still have a link to it...I understand the caution but one of my pet peeves is when healthcare professionals make a choice for me [agreement in room] and say I'm not going to bother to give you the bigger picture and the different options because I think this one is best for you.* [Hugh, pump user]

HCPs considered an SN intervention especially useful for patients who were young adults going through transition or any patient experiencing loneliness. However, patients felt that they themselves would benefit from further support, no matter their circumstances, but according to when they needed it and on their own terms.

Stage 2 Results: Intervention Strategy Selection

Relevant Intervention Functions for Pump Users

Capability

Psychological capabilities were identified in the behavioral diagnosis, and using the intervention mapping matrix, (Figure 3) the following intervention functions were identified: Enablement (a means to increase capability or reduce barriers for SM through encouragement, practical and emotional support, and access to support and opportunities) and Education (increasing knowledge or understanding, including structured education, access to appropriate information, and instructions for performing pump tasks). Physical capability SM barriers and enablers pinpointed to the intervention functions of Training (imparting physical skills in relation to pump technicalities) and Enablement (as described above).

Motivation

Motivational factors of SM that were related to reflective reasoning (conscious intentions, decisions, and plans for SM) led to the intervention functions of Education (as described above). Where there were reflective motivational barriers preventing SM because of support not being seen as relevant or an intervention not being credible, then the intervention functions such as Persuasion through communication to introduce positive feelings to stimulate action or assurance of credibility through research were selected. Where an SN intervention enables SM with self-driven priorities, it increases the likelihood that users will be willing to commit time and that the time they commit will be well spent and valued. Appropriate intervention functions for automatic motivation for SM (emotional responses, desires, and habits) included Persuasion, Environmental restructuring (changing the physical or social context), Modeling (providing an example for people to emulate/aspire to), and Enablement.

Opportunity

Social and physical opportunity to access both emotional and practical support, especially in relation to the specificities and mechanics of a new health technology, was identified in the behavioral diagnosis, and this could be addressed by an SM support web-based intervention. These needs were described in terms of unconventional and flexible ways to self-manage, such as 24/7 access and web-based sources of education, peer support, and information. However, access to any support or resources had to be on the participants' terms, in line with personal needs and life demands, especially in response to concerns over uninvited sharing of SM strategies from others. This was linked with the intervention functions of Enablement and Environmental restructuring (providing access to support, information, and opportunities). Enablement intervention functions were identified to address physical opportunity barriers, such as lack of time to attend or access the clinic or other resources in relation to sourcing support that is physically closer to the individual.

Relevant Intervention Functions for Health Care Professionals

Capability

HCPs said that they believe in prioritizing the wider well-being of their patients and want to support SM, but although they were clear about the medical outcomes they must focus on in their professional role, the remit of SM support they should provide was unclear. HCPs voiced concerns over their lack of confidence, ability, or desire to offer SM support. This is where strategic intervention functions of Training, Enablement, and Education benefit, to instruct HCPs on how to facilitate signposting to an intervention, enable behavioral practice, provide verbal persuasion about capabilities, and educate about the importance of SM support.

Motivation

When it came to Reflective motivational factors, it is evident that buy-in is needed. Coercion (changing conscious evaluations of the SN approach for SM), Education (increasing knowledge or understanding of the importance of social support for their patients), Persuasion (using communication to stimulate action), and Incentivization (creating an expectation of reward—that patients will benefit from the access to SM support) were deemed as appropriate intervention functions, whereas Enablement, Environmental restructuring, and Modeling (comparisons with other clinics) were identified for automatic motivational factors.

Opportunity

Both physical and social opportunity pinpointed to Enablement and Environmental restructuring (provision of physical opportunities and socially acceptable environments to provide SM support) as necessary intervention functions (see Multimedia Appendix 2).

Stage 3 Results: Selection of Specific Behavior Change Techniques

The BCTs identified as likely to benefit the intervention are shown in Multimedia Appendices 1 and 2, with the distinction made between the potentially active ingredients of an intervention ("Reflective" BCTs), which would need to be contained within the SN tool, and the delivery of the intervention ("Strategic" processes), which would need to be integrated into the intervention implementation plan, demonstrated in Table 2. Table 2 also describes where or with whom the intervention reflective and strategic processes are to be implemented. These intervention "ingredients" are categorized as in terms of being; the role of the facilitator, an intervention function, within the study protocol, as an invitation to take part, within a site initiation visit, in training, or in future research. This addresses the varying needs and expectations of insulin pump users and HCPs, and this suggests how a web-based intervention designed to enable SM support can attend to these.

Table 2. Identified behavior change techniques of intervention (reflective or strategy processes).

Identified needs (behavior change techniques)	SN[a] intervention ingredients	Where change is to be implemented/delivered
Reflective processes		
Goal setting (behavior)	Agreement to attend a preferred activity identified in the intervention	Role of facilitator
Problem solving	An SN tool maps the participants social support network and examines whether the participant would like this to change at all. The intervention also inquires about their personal needs and preferences and then offers opportunities in their local community to address these needs. A discussion is then undertaken about how to access these, as well as barriers and facilitators	Intervention function and Role of facilitator
Feedback on behavior	The facilitator follows-up with the participants and discusses and informs them of how their circles have changed and what activities have been taken up	Role of facilitator
Social support (unspecified)	GENIE[b] facilitates discussion around who offers them social support in relation to their condition and allows facilitation/gives information about further personalized social support, that is, peer support groups, and asks who may help them participate in chosen activities	Intervention function and Role of facilitator
Social support (practical)	Discuss the practical support required, received, and desired from the participant and facilitate discussion over whether any changes are required and how to undertake these changes or discuss how existing members of the participant's SN can help them physically access groups	Intervention function and Role of facilitator
Social support (emotional)	Discuss the emotional support required, received, and desired from the participant and facilitate discussion over whether any changes are required and how to undertake these changes or discuss how existing members of the participant's SN can help them feel emotionally able to access groups	Intervention function and Role of facilitator
Instruction on how to perform a behavior	If a person wants to attend a course or education session, then GENIE can facilitate access to this, or if a person wants to learn from peers, then GENIE can point them in the direction of a peer support group	Intervention function
Prompts/cues	GENIE comprises concentric circles, which prompt the participant to prioritize certain SN members over others. GENIE then asks 13 preference questions to prompt the user regarding the user's preferred activities to support SM[c]. Participants are then followed up by a facilitator after 2 weeks	Intervention function
Comparative imagining of future outcomes	Prompt the participant to imagine and compare likely or possible outcomes following attending versus not attending particular groups or activities in which they took part	Role of facilitator
Reduce negative emotions	The facilitator advises to use members of the current social support network to reduce anxiety about attending groups	Role of facilitator
Conserving mental resources	The facilitator advises to utilize the social support network or access peer support groups to share the burden of diabetes or to find someone to troubleshoot with	Role of facilitator
Restructuring the physical environment	Enabling access to groups and information that can help them engage in SM	Intervention function
Restructuring the social environment	Enabling access to and restructuring groups, information, and support that can help them engage in SM	Intervention function
Framing/reframing	The facilitator reassures participant that it is okay to ask for help or support from others regarding SM and that others can offer practical tips	Role of facilitator
Focus on past success	The facilitator enquires about activities they used to do and whether the network members can assist their attendance at activities in which they are interested	Role of facilitator
Strategy processes		
Action planning	Steps would need to be taken to support each clinic to implement the intervention and identify pathways	Protocol and Site initiation visit

Identified needs (behavior change techniques)	SN[a] intervention ingredients	Where change is to be implemented/delivered
Review behavior goals	The clinic would need to be reviewed to identify whether further support is required to implement the intervention	Protocol and Continuous communication from the research team
Behavioral contract	The clinic would need to sign a contract to identify what they expect from the intervention and what support they require	Protocol and Agreements
Commitment	The clinic would need to make SM support a priority and normalized within the clinic setting and be committed to offering SM support	Site initiation visit and Continuous communication from the research team
Instructions on how to perform a behavior	Facilitators of GENIE receive training on how to deliver GENIE. The tool currently comes with a training program	Protocol and Training
Behavioral experiments	Pilot study intervention with clinics to demonstrate intervention benefits in this patient group/context	Future research
Demonstration of the behavior	Facilitators of GENIE receive training on how to deliver GENIE. The tool currently comes with a training program	Protocol and Training
Information about others' approval	Share experiences from other clinics/areas using the tool	Future research and Site initiation visit
Behavioral practice/rehearsal	Facilitators of GENIE receive training on how to deliver GENIE. The tool currently comes with a training program	Protocol and Training
Credible source	Buy-in from each area it is applied to is important for implementation. Participants (and HCPs[d]) are assured that GENIE has risen out of former research and that everything put on GENIE is checked	Invitation to take part; Protocol; Site initiation visit; and Training
Comparative imagining of future outcomes	Prompt the clinic to imagine and compare likely or possible outcomes following implementation of GENIE	Invitation to take part; Protocol; Site initiation visit; and Training
Restructuring the physical environment	Enabling access to SM support and information that can help patients engage in SM	Invitation to take part; Protocol; Site initiation visit; and Training
Restructuring the social environment	Enabling physical access to groups and information and support that can help patients engage in SM	Invitation to take part; Protocol; Site initiation visit; and Training
Framing/reframing	Draw attention to research suggesting that SM support can provide clinical benefits and reduced health utilization. SM support could therefore increase clinic time available rather than decrease clinic time.	Protocol; Site initiation visit; and Training
Incompatible beliefs	Draw attention to how restricting the provision of SM support is in contrast with national guidance (National Health Service England) which promotes SM support.	Invitation to take part; Protocol; Site initiation visit; and Training

[a]SN: social network.

[b]GENIE: Generating Engagement in Networks InvolvEment.

[c]SM: self-management.

[d]HCP: health care professional.

Discussion

Principal Findings

This study provides a model for supporting people who are incorporating a health technology that is new to them (such as an insulin pump), through consideration of key stakeholders' needs in developing an intervention that aims to provide SM support. This study puts the most important needs at the forefront (stakeholders' needs), providing evidence of the active components required in a translational web-based intervention. In this instance, the physicality of the pump (the new device) impacts the users' experience of SM, and the technicalities while using an advanced technology exacerbate SM needs. The specificity of the insulin pump changes people's priorities, as it impacts their day-to-day experiences and identity. Pump therapy means that users have a renewed need for HCPs, akin to the diagnosis of diabetes, but this need subsides. The pump requires access to a particular network of people for specific troubleshooting needs.

With this in mind, we identified that a long-term condition such as diabetes requires an array of SM approaches and the ability to master these. Utilizing a new health technology or device requires specific skills, understanding, confidence, motivation, and opportunity. The behavioral analysis used here signposted the necessary components of an intervention to support SM. For example, there lies a potential conflict for the person living with T1D, where "good" management takes considerable effort, and this can create a friction between freedom and clinical targets of blood glucose control or the opportunity for tighter control without sacrificing freedom. The question arises as to whether this extra attention will actually improve the quality and length of life. The current SM support options offered to

people with a long-term condition such as T1D can incite questions over whether life will be less or more enjoyable if they take part in them, for example, using 5 days of annual leave to attend an NHS structured–education class, not knowing whether this education class will actually be useful. There is a trade-off to be made. If we want to intervene, then we must consider these factors. We found that ultimate behavior change in the SM of diabetes and use of a health technology requires support and resources, the availability of which is personal/specific to the individual and varies according to time and life circumstances. Specific social support can take away some of the work of SM, as well as the isolation, providing shared learnings and practical tips, but limitations include fear of judgment from others and exposure to off-putting self-pity from peers.

We found that an intervention would be more successfully implemented if there were opportunities to access SM support and motivation from pump users by access to relevant disease/technology-specific resources and interests. For example, social opportunity needs to be addressed when HCPs do not entrust pump users with the ability to make their own choices or access nonclinical resources or when HCPs lack psychological capability and/or physical opportunity, with HCPs (even if willing) often not having the confidence or capacity in their clinic to amend or enhance their routine or psychosocial care and questioning whether and where SM fits into their role. The recommendations provided here for delivering training to HCPs to facilitate signposting to holistic SM support, enabling behavioral practice, providing verbal persuasion about capabilities, and educating about the importance of SM echo those given in the DAWN2 study [52,53]. Guidelines within the intervention could give assurances to HCPs about what they are signposting to. However, some HCPs' "if-it-isn't-broke-don't-fix-it" attitude highlighted that although NHS England is pushing for more SM support, it is not reaching or convincing to all clinicians on the ground. Fisher et al [54] suggest a clear 3-step framework for diabetes HCPs to support behavior change. The first step requires clinicians to shift their mind-set, moving from a hierarchal model to a more collaborative model, reorienting from information giving to nuances of patient-driven needs.

HCPs can be seen as gatekeepers or blockers of the provision of SM support necessary to manage a complex condition such as diabetes. There was little doubt among HCPs, even those with general concerns, that particular groups of patients would greatly benefit from being signposted to further support that the clinic did not provide. However, the discrepancies show contrasting beliefs between patients and HCPs, where patients themselves considered this access beneficial in a variety of ways, especially in terms of managing fitness activities, general practical advice, or emotional support. Credibility and likelihood of the effectiveness of an intervention are, unsurprisingly, important for both users and those who guard access (clinicians), particularly for clinicians who offer patients the opportunity to participate. Priorities vary depending on perspective, and understanding both perspectives at this stage can inform intervention design and how to determine and ensure credibility.

The SN intervention proposed here (GENIE) is structured around facilitating networks and collective, tailored forms of support through the building of dedicated resources in a database. Whether targeting particular groups or long-term conditions as a whole, a web-based SN tool can accommodate multiple SM needs [55]. However, limitations are evident where access to resources is only as good as the resources that are already in place locally. An SN intervention such as this would also benefit people by addressing the identified need to register collective interests and initiate peer support. For example, having the facility for people to "register their interests" in attending or creating groups in their local area or the ability to connect with others in their local area via the intervention platform.

The behavioral assessment of people with diabetes and HCPs draws parallels with past research. For example, Mulvaney et al's [36] review concluded that SM interventions in diabetes should integrate technology compounded with human contact for clinical support, as well as motivation and support to change behavior for SM (eg, goal setting and problem solving). In addition, the American Diabetes Association [56] considers behavioral elements such as problem solving, decision making, and providing access to electronic health tools as vital to support SM. A focused SN intervention with integral guided facilitation in place is likely to be sensitive to these needs, combined with participant follow-up from the facilitator. A facilitator also has the potential to provide a favorable supportive element to personalized goals in light of findings that the provision of human support was advantageous in other electronic health interventions [57].

People who are empowered and skilled to self-manage their diabetes have improved health outcomes [6,58,59]; therefore, appropriate and tailored access, as opposed to a one-size-fits-all model, is likely to support improved SM. HCPs need to accept patient priorities and means of information and advocacy [60,61] and understanding the importance of experiential evidence. Some noted factors of success in web-based interventions and acceptability have been the focus on psychosocial experiences, feelings of confidence and reduced fear, the availability outside of clinic hours, up-to-date evidence-based guidance, and access to both peer-generated and professional advice [55,58,59,61-65]. However, understanding the barriers preventing HCPs from supporting SM is fundamental too [61,65]. This comprehensive behavioral analysis provides a complete feedback loop for a web-based intervention, which is better equipped to facilitate ongoing SM, considering the needs and strategies for both sets of stakeholders, and determine how, when, and why SM support interventions can be best utilized.

Strengths and Limitations

The use of focus groups in this study allows an in-depth discussion and understanding of the collective experiences of SM and of patients' and HCPs' views, which would be impossible to explore using quantitative methods, and the use of the BCW and TDF-driven interview scripts provides a well-tested, evidence-based guideline and framework. For example, it has been noted that the automatic addition to the reflective process of motivation to enact behavior on the part

of HCPs is often overlooked and is important to enhance the behavioral approaches to implementation [66]. The use of theory-driven intervention development signifies areas of key importance to intervention implementation, both behavioral and reflective needs and contextual factors for implementation, and it is a key process to follow. It sensitizes the research to future intervention needs and considerations across different localities. However, although the proposed BCTs were carefully considered in response to stakeholder needs, these have not been developed in consultation with stakeholders, which future work needs to address to verify BCT feasibility. The recruitment of pump users from various clinics and the involvement of clinics in different settings were important elements of the expected variability among local health systems. Although the participants recruited represented a variety of ages and sex, education attainment, and parenthood, the clinicians represented a good and balanced spectrum of the kinds of professionals working in insulin pump clinics. However, the limitations to the study were that the recruited patients were more likely to be those that were particularly open to discussing personal elements of their diabetes management and willing to sit, in their own time, among a group of peers with the same condition, and therefore would not necessarily represent a number of people living with T1D.

Future Research and Conclusions

Technology is a means to deal with diabetes, and it opens new ways to manage the condition, but it takes time to master; therefore, appropriate support, skills, and information are crucial. People with T1D have a uniqueness of knowledge about their own body, which challenges professional dominance and creates an invisible barrier wherein despite HCPs possessing sound medical knowledge, they are unsure of what and when to share with their patients. HCPs can be gatekeepers for improving SM or for facilitating access to SM support. They are limited by time constraints and fear of professional responsibility. However, a web-based tool that is person based, appropriate, accessible, and adaptive to local needs, along with a strategic (and theoretically informed) approach, can provide a powerful tool for SM support, which can vastly enhance the support already being provided by HCPs [55,63]. This paper has strived to demonstrate the development of such an intervention. The study is particularly timely in that it coincides with The NHS Long Term Plan from NHS England, January 2019, which promises to expand the provision of digital SM support tools [67]. In addition, there has been a recent drive for the integration of psychosocial support into routine diabetes care [18,21], and this study provides an initial engagement with the factors that would impact how psychosocial support is taken up with HCPs and the priorities for patients. The next phase of development is to integrate these findings into strategic intervention implementation criteria for supporting people to engage in SM with a new device and technology such as an insulin pump.

Acknowledgments

The authors wish to specially thank all the participants of the study: people living with diabetes, specialist HCPs and clinics, and the diabetes voluntary organizations and peer support groups that kindly disseminated the advertising for the focus groups. The authors are grateful for the opportunity to undertake this doctoral research through funding from The Health Foundation, and this study was supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care (Wessex). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Matrix of links among the Capability, Opportunity, Motivation-Behavior model, Theoretical Domains Framework domains, intervention functions, and behavior change techniques for pump users.
[DOCX File , 25 KB - jmir_v22i5e13980_app1.docx]

Multimedia Appendix 2
Matrix of links among the Capability, Opportunity, Motivation-Behavior model, Theoretical Domains Framework domains, intervention functions, and behavior change techniques for health care professionals.
[DOCX File , 18 KB - jmir_v22i5e13980_app2.docx]

References

1. Patterson CC, Harjutsalo V, Rosenbauer J, Neu A, Cinek O, Skrivarhaug T, et al. Trends and cyclical variation in the incidence of childhood type 1 diabetes in 26 European centres in the 25 year period 1989-2013: a multicentre prospective registration study. Diabetologia 2019 Mar;62(3):408-417. [doi: 10.1007/s00125-018-4763-3] [Medline: 30483858]

2. Foster NC, Beck RW, Miller KM, Clements MA, Rickels MR, DiMeglio LA, et al. State of type 1 diabetes management and outcomes from the T1D exchange in 2016-2018. Diabetes Technol Ther 2019 Feb;21(2):66-72. [doi: 10.1089/dia.2018.0384] [Medline: 30657336]
3. NHS Digital.: NHS Digital; 2017. National Diabetes Audit Report 1 Care Processes and Treatment Targets 2016-17 URL: https://digital.nhs.uk/data-and-information/publications/statistical/national-diabetes-audit/national-diabetes-audit-report-1-care-processes-and-treatment-targets-2016-17 [accessed 2019-03-11]
4. McBrien KA, Manns BJ, Hemmelgarn BR, Weaver R, Edwards AL, Ivers N, et al. The association between sociodemographic and clinical characteristics and poor glycaemic control: a longitudinal cohort study. Diabet Med 2016 Nov;33(11):1499-1507. [doi: 10.1111/dme.13023] [Medline: 26526088]
5. El-Gayar O, Timsina P, Nawar N, Eid W. A systematic review of IT for diabetes self-management: are we there yet? Int J Med Inform 2013 Aug;82(8):637-652. [doi: 10.1016/j.ijmedinf.2013.05.006] [Medline: 23792137]
6. Chatterjee S, Davies MJ, Heller S, Speight J, Snoek FJ, Khunti K. Diabetes structured self-management education programmes: a narrative review and current innovations. Lancet Diabetes Endocrinol 2018 Feb;6(2):130-142. [doi: 10.1016/S2213-8587(17)30239-5] [Medline: 28970034]
7. Campbell F, Lawton J, Rankin D, Clowes M, Coates E, Heller S, et al. Follow-Up Support for Effective type 1 Diabetes self-management (The FUSED Model): a systematic review and meta-ethnography of the barriers, facilitators and recommendations for sustaining self-management skills after attending a structured education programme. BMC Health Serv Res 2018 Dec 27;18(1):898 [FREE Full text] [doi: 10.1186/s12913-018-3655-z] [Medline: 30482202]
8. Umpierrez GE, Klonoff DC. Diabetes Technology Update: use of insulin pumps and continuous glucose monitoring in the hospital. Diabetes Care 2018 Aug;41(8):1579-1589 [FREE Full text] [doi: 10.2337/dci18-0002] [Medline: 29936424]
9. Nicolucci A, Burns KK, Holt RI, Lucisano G, Skovlund SE, Kokoszka A, et al. Correlates of psychological outcomes in people with diabetes: results from the second Diabetes Attitudes, Wishes and Needs (DAWN2) study. Diabet Med 2016 Sep;33(9):1194-1203. [doi: 10.1111/dme.13178] [Medline: 27353119]
10. Reidy C, Bracher M, Foster C, Vassilev I, Rogers A. The process of incorporating insulin pumps into the everyday lives of people with Type 1 diabetes: a critical interpretive synthesis. Health Expect 2018 Aug;21(4):714-729 [FREE Full text] [doi: 10.1111/hex.12666] [Medline: 29418050]
11. Prahalad P, Tanenbaum M, Hood K, Maahs DM. Diabetes technology: improving care, improving patient-reported outcomes and preventing complications in young people with Type 1 diabetes. Diabet Med 2018 May;35(4):419-429. [doi: 10.1111/dme.13588] [Medline: 29356074]
12. 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 Dec 10;6(4):e40 [FREE Full text] [doi: 10.2196/jmir.6.4.e40] [Medline: 15631964]
13. Hood KK, Hilliard M, Piatt G, Ievers-Landis CE. Effective strategies for encouraging behavior change in people with diabetes. Diabetes Manag (Lond) 2015;5(6):499-510 [FREE Full text] [doi: 10.2217/dmt.15.43] [Medline: 30100925]
14. Diabetes UK. Leicester: NHS Diabetes and Diabetes UK; 2010. Emotional and Psychological Support and Care in Diabetes URL: https://www.diabetes.org.uk/professionals/position-statements-reports/diagnosis-ongoing-management-monitoring/emotional-and-psychological-support-and-care-in-diabetes [accessed 2019-03-11]
15. Diabetes UK. Diabetes UK. London: Diabetes UK; 2008. Minding the Gap: The Provision of Psychological Support and Care for People With Diabetes in the UK URL: https://www.diabetes.org.uk/resources-s3/2017-11/minding_the_gap_psychological_report.pdf [accessed 2019-03-11]
16. Diabetes UK. Diabetes UK. 2016. Position Statement: Emotional and Psychological Support for People With Diabetes URL: https://www.diabetes.org.uk/resources-s3/2017-10/Revised%20Emotional%20and%20psychological%20support_DUK%20position%20statement_%28For%20web%20-%20without%20info%20prescrip%20text%29%20-%20Revised%20reference%203%2010%202017.pdf [accessed 2019-03-11]
17. Jones A, Vallis M, Pouwer F. If it does not significantly change HbA1c levels why should we waste time on it? A plea for the prioritization of psychological well-being in people with diabetes. Diabet Med 2015 Mar;32(2):155-163. [doi: 10.1111/dme.12620] [Medline: 25354315]
18. Joensen L, Fisher L, Skinner T, Doherty Y, Willaing I. Integrating psychosocial support into routine diabetes care: perspectives from participants at the Self-Management Alliance meeting 2016. Diabet Med 2019 Jul;36(7):847-853. [doi: 10.1111/dme.13836] [Medline: 30315608]
19. Rogers A, Vassilev I, Sanders C, Kirk S, Chew-Graham C, Kennedy A, et al. Social networks, work and network-based resources for the management of long-term conditions: a framework and study protocol for developing self-care support. Implement Sci 2011 May 29;6:56 [FREE Full text] [doi: 10.1186/1748-5908-6-56] [Medline: 21619695]
20. Rogers A, Brooks H, Vassilev I, Kennedy A, Blickem C, Reeves D. Why less may be more: a mixed methods study of the work and relatedness of 'weak ties' in supporting long-term condition self-management. Implement Sci 2014 Mar 13;9:19 [FREE Full text] [doi: 10.1186/1748-5908-9-19] [Medline: 24524253]
21. Gallant MP. The influence of social support on chronic illness self-management: a review and directions for research. Health Educ Behav 2003 May;30(2):170-195. [doi: 10.1177/1090198102251030] [Medline: 12693522]

22. Wysocki T, Greco P. Social support and diabetes management in childhood and adolescence: influence of parents and friends. Curr Diab Rep 2006 May;6(2):117-122. [doi: 10.1007/s11892-006-0022-y] [Medline: 16542622]
23. Rosland AM, Kieffer E, Israel B, Cofield M, Palmisano G, Sinco B, et al. When is social support important? The association of family support and professional support with specific diabetes self-management behaviors. J Gen Intern Med 2008 Dec;23(12):1992-1999 [FREE Full text] [doi: 10.1007/s11606-008-0814-7] [Medline: 18855075]
24. Schiøtz ML, Bøgelund M, Almdal T, Jensen BB, Willaing I. Social support and self-management behaviour among patients with Type 2 diabetes. Diabet Med 2012 May;29(5):654-661. [doi: 10.1111/j.1464-5491.2011.03485.x] [Medline: 21992512]
25. Joensen LE, Almdal TP, Willaing I. Type 1 diabetes and living without a partner: psychological and social aspects, self-management behaviour, and glycaemic control. Diabetes Res Clin Pract 2013 Oct;101(3):278-285. [doi: 10.1016/j.diabres.2013.07.001] [Medline: 23891305]
26. 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;9(6):e98340 [FREE Full text] [doi: 10.1371/journal.pone.0098340] [Medline: 24887107]
27. Allen C, Vassilev I, Kennedy A, Rogers A. Long-term condition self-management support in online communities: a meta-synthesis of qualitative papers. J Med Internet Res 2016 Mar 10;18(3):e61 [FREE Full text] [doi: 10.2196/jmir.5260] [Medline: 26965990]
28. Blakeman T, Blickem C, Kennedy A, Reeves D, Bower P, Gaffney H, et al. Effect of information and telephone-guided access to community support for people with chronic kidney disease: randomised controlled trial. PLoS One 2014;9(10):e109135 [FREE Full text] [doi: 10.1371/journal.pone.0109135] [Medline: 25330169]
29. Kennedy A, Vassilev I, James E, Rogers A. Implementing a social network intervention designed to enhance and diversify support for people with long-term conditions. A qualitative study. Implement Sci 2016 Mar 29;11:27 [FREE Full text] [doi: 10.1186/s13012-016-0384-8] [Medline: 26926837]
30. Gonzalez JS, Tanenbaum ML, Commissariat PV. Psychosocial factors in medication adherence and diabetes self-management: Implications for research and practice. Am Psychol 2016 Oct;71(7):539-551 [FREE Full text] [doi: 10.1037/a0040388] [Medline: 27690483]
31. Funnell MM. The Diabetes Attitudes, Wishes, and Needs (DAWN) Study. Clin Diabetes 2006;24(4):154-155. [doi: 10.2337/diaclin.24.4.154]
32. Barnard KD, Peyrot M, Holt RI. Psychosocial support for people with diabetes: past, present and future. Diabet Med 2012 Dec;29(11):1358-1360. [doi: 10.1111/j.1464-5491.2012.03727.x] [Medline: 22681530]
33. Reddy M, Rilstone S, Cooper P, Oliver NS. Type 1 diabetes in adults: supporting self management. Br Med J 2016 Mar 10;352:i998. [doi: 10.1136/bmj.i998] [Medline: 26965473]
34. Campbell NC, Murray E, Darbyshire J, Emery J, Farmer A, Griffiths F, et al. Designing and evaluating complex interventions to improve health care. Br Med J 2007 Mar 3;334(7591):455-459 [FREE Full text] [doi: 10.1136/bmj.39108.379965.BE] [Medline: 17332585]
35. Pilkington K, Loef M, Polley M. Searching for real-world effectiveness of health care innovations: scoping study of social prescribing for diabetes. J Med Internet Res 2017 Feb 2;19(2):e20 [FREE Full text] [doi: 10.2196/jmir.6431] [Medline: 28153817]
36. Mulvaney SA, Ritterband LM, Bosslet L. Mobile intervention design in diabetes: review and recommendations. Curr Diab Rep 2011 Dec;11(6):486-493. [doi: 10.1007/s11892-011-0230-y] [Medline: 21960031]
37. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. Br Med J 2008 Oct 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
38. Michie S, Atkins LR, West R. The Behaviour Change Wheel: A Guide To Designing Interventions. London: Silverback Publishing; 2014.
39. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 May 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
40. Murphy ME, Byrne M, Zarabzadeh A, Corrigan D, Fahey T, Smith SM. Development of a complex intervention to promote appropriate prescribing and medication intensification in poorly controlled type 2 diabetes mellitus in Irish general practice. Implement Sci 2017 Sep 16;12(1):115 [FREE Full text] [doi: 10.1186/s13012-017-0647-z] [Medline: 28915897]
41. Green J, Thorogood N. Qualitative Methods for Health Research. London: Sage; 2009.
42. Wilkinson S. Focus groups in health research: exploring the meanings of health and illness. J Health Psychol 1998 Jul;3(3):329-348. [doi: 10.1177/135910539800300304] [Medline: 22021395]
43. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
44. Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. Implement Sci 2012 May 24;7:37 [FREE Full text] [doi: 10.1186/1748-5908-7-37] [Medline: 22530986]

45. Michie S, Johnston M, Abraham C, Lawton R, Parker D, Walker A, Psychological Theory Group. Making psychological theory useful for implementing evidence based practice: a consensus approach. Qual Saf Health Care 2005 Mar;14(1):26-33 [FREE Full text] [doi: 10.1136/qshc.2004.011155] [Medline: 15692000]
46. Atkins L, Francis J, Islam R, O'Connor D, Patey A, Ivers N, et al. A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. Implement Sci 2017 Jun 21;12(1):77 [FREE Full text] [doi: 10.1186/s13012-017-0605-9] [Medline: 28637486]
47. Curtis KE, Lahiri S, Brown KE. Targeting parents for childhood weight management: development of a theory-driven and user-centered healthy eating app. JMIR Mhealth Uhealth 2015 Jul 18;3(2):e69 [FREE Full text] [doi: 10.2196/mhealth.3857] [Medline: 26088692]
48. Phillips CJ, Marshall AP, Chaves NJ, Jankelowitz SK, Lin IB, Loy CT, et al. Experiences of using the Theoretical Domains Framework across diverse clinical environments: a qualitative study. J Multidiscip Healthc 2015;8:139-146 [FREE Full text] [doi: 10.2147/JMDH.S78458] [Medline: 25834455]
49. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The Behavior Change Technique Taxonomy (v1) of 93 Hierarchically Clustered Techniques: Building an International Consensus for the Reporting of Behavior Change Interventions. Annals of Behavioral Medicine 2013;46(1):81-95. [doi: 10.1007/s12160-013-9486-6]
50. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Dec;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
51. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
52. Holt RI, Nicolucci A, Burns KK, Lucisano G, Skovlund SE, Forbes A, et al. Correlates of psychological care strategies for people with diabetes in the second Diabetes Attitudes, Wishes and Needs (DAWN2) study. Diabet Med 2016 Sep;33(9):1174-1183. [doi: 10.1111/dme.13109] [Medline: 26939906]
53. Byrne JL, Davies MJ, Willaing I, Holt RI, Carey ME, Daly H, et al. Deficiencies in postgraduate training for healthcare professionals who provide diabetes education and support: results from the Diabetes Attitudes, Wishes and Needs (DAWN2) study. Diabet Med 2017 Aug;34(8):1074-1083. [doi: 10.1111/dme.13334] [Medline: 28195662]
54. Fisher L, Polonsky WH, Hessler D, Potter MB. A practical framework for encouraging and supporting positive behaviour change in diabetes. Diabet Med 2017 Dec;34(12):1658-1666 [FREE Full text] [doi: 10.1111/dme.13414] [Medline: 28636745]
55. Dobson R, Whittaker R, Murphy R, Khanolkar M, Miller S, Naylor J, et al. The use of mobile health to deliver self-management support to young people with type 1 diabetes: a cross-sectional survey. JMIR Diabetes 2017 Mar 15;2(1):e4 [FREE Full text] [doi: 10.2196/diabetes.7221] [Medline: 30291057]
56. Beck J, Greenwood DA, Blanton L, Bollinger ST, Butcher MK, Condon JE, 2017 Standards Revision Task Force. 2017 national standards for diabetes self-management education and support. Diabetes Care 2017 Oct;40(10):1409-1419. [doi: 10.2337/dci17-0025] [Medline: 28754780]
57. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar 10;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
58. Oser SM, Stuckey HL, Parascando JA, McGinley EL, Berg A, Oser TK. Glycated hemoglobin differences among blog-reading adults with type 1 diabetes compared with those who do not read blogs: cross-sectional study. JMIR Diabetes 2019 May 2;4(2):e13634 [FREE Full text] [doi: 10.2196/13634] [Medline: 30938693]
59. Litchman ML, Edelman LS, Donaldson GW. Effect of diabetes online community engagement on health indicators: cross-sectional study. JMIR Diabetes 2018 May 24;3(2):e8 [FREE Full text] [doi: 10.2196/diabetes.8603] [Medline: 30291079]
60. Mazanderani F, O'Neill B, Powell J. 'People power' or 'pester power'? YouTube as a forum for the generation of evidence and patient advocacy. Patient Educ Couns 2013 Dec;93(3):420-425 [FREE Full text] [doi: 10.1016/j.pec.2013.06.006] [Medline: 23830239]
61. Reidy C, Klonoff DC, Barnard-Kelly KD. Supporting good intentions with good evidence: how to increase the benefits of diabetes social media. J Diabetes Sci Technol 2019 Oct;13(5):974-978. [doi: 10.1177/1932296819850187] [Medline: 31096768]
62. Oser TK, Oser SM, McGinley EL, Stuckey HL. A novel approach to identifying barriers and facilitators in raising a child with type 1 diabetes: qualitative analysis of caregiver blogs. JMIR Diabetes 2017 Oct 26;2(2):e27 [FREE Full text] [doi: 10.2196/diabetes.8966] [Medline: 30291073]
63. Oser TK, Minnehan KA, Wong G, Parascando J, McGinley E, Radico J, et al. Using social media to broaden understanding of the barriers and facilitators to exercise in adults with type 1 diabetes. J Diabetes Sci Technol 2019 May;13(3):457-465. [doi: 10.1177/1932296819835787] [Medline: 30862185]
64. Prakasam G, Rees C, Lyden M, Parkin CG. Use of a novel smartphone-based diabetes management system improved feelings of confidence and safety and reduced hypoglycemia fear among parents/caregivers of children/adolescents with type 1 diabetes. J Diabetes Sci Technol 2017 Jan;11(1):182-183 [FREE Full text] [doi: 10.1177/1932296816650901] [Medline: 27207891]

65. Ventola CL. Social media and health care professionals: benefits, risks, and best practices. Pharm Ther 2014 Jul;39(7):491-520 [FREE Full text] [Medline: 25083128]
66. Potthoff S, Presseau J, Sniehotta FF, Breckons M, Rylance A, Avery L. Exploring the role of competing demands and routines during the implementation of a self-management tool for type 2 diabetes: a theory-based qualitative interview study. BMC Med Inform Decis Mak 2019 Jan 24;19(1):23 [FREE Full text] [doi: 10.1186/s12911-019-0744-9] [Medline: 30678684]
67. NHS England. NHS England.: NHS England; 2019. NHS Long Term Plan URL: https://www.longtermplan.nhs.uk/ [accessed 2019-03-11]

Abbreviations

APPEASE: Acceptability, Practicability, Effectiveness/Cost-effectiveness, Affordability, Safety, Equity
BCT: behavior change technique
BCW: Behavior Change Wheel
COM-B: Capability, Opportunity, Motivation-Behavior
DAWN2: Diabetes Attitudes, Wishes, and Needs second study
GENIE: Generating Engagement in Networks InvolvEment
HCP: health care professional
MDI: multiple daily injections
NHS: National Health Service
NIHR: National Institute for Health Research
SM: self-management
SN: social network
T1D: type 1 diabetes
TDF: Theoretical Domains Framework

Edited by G Eysenbach; submitted 11.03.19; peer-reviewed by G Forbes, S Potthoff, S Ge, S Oser; comments to author 13.09.19; revised version received 08.11.19; accepted 16.12.19; published 01.05.20.

Please cite as:
Reidy C, Foster C, Rogers A
A Facilitated Web-Based Self-Management Tool for People With Type 1 Diabetes Using an Insulin Pump: Intervention Development Using the Behavior Change Wheel and Theoretical Domains Framework
J Med Internet Res 2020;22(5):e13980
URL: https://www.jmir.org/2020/5/e13980
doi:10.2196/13980
PMID:32356776

Original Paper

Web-Based Intervention Effects on Mild Cognitive Impairment Based on Apolipoprotein E Genotype: Quasi-Experimental Study

Anthoula C Tsolaki[1,2], MD; Magda Tsolaki[3], MD, PhD; Niki Pandria[1], MSc; Eftychia Lazarou[3], MSc; Olymbia Gkatzima[4], MSc; Vasiliki Zilidou[1], MSc; Maria Karagianni[1], MSc; Zafiroula Iakovidou-Kritsi[5], PhD; Vasilios K Kimiskidis[6], MD, PhD; Panagiotis D Bamidis[1], MSc, PhD

[1]Medical Physics Laboratory, School of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece

[2]Department of Neurology, Agios Pavlos General Hospital, Thessaloniki, Greece

[3]1st Department of Neurology, American Hellenic Educational Progressive Association Hospital, Aristotle University of Thessaloniki, Thessaloniki, Greece

[4]Panhellenic Institute of Neurodegenerative Diseases, Thessaloniki, Greece

[5]Laboratory of Medical Biology-Genetics Department, School of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece

[6]Laboratory of Clinical Neurophysiology, American Hellenic Educational Progressive Association Hospital, Aristotle University of Thessaloniki, Thessaloniki, Greece

Corresponding Author:
Panagiotis D Bamidis, MSc, PhD
Medical Physics Laboratory
School of Medicine
Aristotle University of Thessaloniki
PO Box 376
School of Medicine, Aristotle University of Thessaloniki
Thessaloniki
Greece
Phone: 30 2310999310
Email: pdbamidis@gmail.com

Abstract

Background: Apolipoprotein E (APOE) ε4 allele is a major genetic risk factor for Alzheimer disease and mild cognitive impairment (MCI). Computer-based training programs can improve cognitive performance in elderly populations. However, the effects of computer-based interventions on MCI APOE ε4 carriers have never been studied before.

Objective: The effects of different web-based interventions and the APOE isoform-specific differences in training outcomes are investigated.

Methods: Using a quasi-experimental study design, 202 participants with MCI aged 60 years and older took part in three different intervention programs (physical and cognitive [Long-Lasting Memories, or LLM], cognitive [Active Control, or AC], or physical intervention [Physical Training Control, or PTC]) via an innovative information and communication technologies exergaming platform. Participants in each interventional group were subdivided into APOE ε4 carriers and non–APOE ε4 carriers. All participants underwent an extensive neuropsychological evaluation before and after the training, blood tests, and brain imaging.

Results: All interventions resulted in multiple statistically significant cognitive benefits after the intervention. Verbal learning (California Verbal Learning Test: immediate recall test score—LLM: P=.04; AC: P<.001), working memory (digit span forward and backward test scores—AC: P=.03; PTC: P=.02 and P=.006, respectively), and long-term memory (California Verbal Learning Test: delayed recall test score—LLM: P=.02; AC: P=.002; and PTC: P=.02) were improved. There was no statistically significant difference among the intervention effects. APOE ε4 presence moderates intervention effects as the LLM intervention improved only their task-switching processing speed (Trail Making Test, Part B: P=.03) and the PTC intervention improved only the working memory (digit span backward: P=.03). No significant performance alteration was noted for the APOE ε4+ cognitive AC training group.

Conclusions: None of the applied interventions could be identified as the optimal one; it is suggested, however, that combined cognitive and physical training and physical training via exergaming may be more effective for the high-risk MCI APOE ε4+ subgroup.

(J Med Internet Res 2020;22(5):e14617) doi:10.2196/14617

KEYWORDS

mild cognitive impairment; APOE ε4; computerized training; exergaming; Alzheimer disease

Introduction

Mild Cognitive Impairment

Mild cognitive impairment (MCI) can be defined as the condition between normal aging and dementia [1]. By the time older people meet the criteria for MCI, they have already exhibited measurable cognitive decline, and most of them have also accumulated the neuropathologic hallmarks of Alzheimer disease [2]. Heterogeneous etiology can cause MCI. Due to this heterogeneity, its progression is uncertain; patients may remain stable for years, a few could improve, and others could progress to dementia. Based on recent data, MCI presents a progression rate of 38.7% over 12 to 60 months. However, the progression rate across individual studies is quite variable, ranging from 6% to 39% per year [3].

Even though MCI is regarded as a preclinical stage of Alzheimer disease or other types of dementia, studies report neuronal loss of about 36.5% already at that stage as well as synaptic dysfunction [4]. MCI patients, however, seem to retain sufficient neuroplasticity to benefit from nonpharmacological interventions, which may, in turn, delay the progression to dementia [5]. Since there is currently no other treatment of dementia than the palliative one, research efforts are focused on possible ways that could delay disease onset, such as diet, cognitive, and physical training.

Nonpharmaceutical Interventions

The recent growing interest in investigating interventions capable of ameliorating or delaying aging and neurodegenerative effects has resulted in designing various projects for cognitive or/and physical training of the elderly [6]. There is no consensus about the best combination of training. However, findings suggest that interventions targeting multiple domains may be more effective and even provide a long-term benefit for individuals at risk [7].

Cognitive training is based on the idea that the brain function is modifiable even in old age [6]. It is claimed that cognitive training may contribute to the delay or even prevention of cognitive decline in older adults, although this claim remains controversial [8]. Cognitive improvement after cognitive training is generally associated with both compensatory and restorative mechanisms [9-12].

Physical training seems to promote multiple gains in both physical and cognitive states. Hippocampal neurogenesis [13], decrease of β-amyloid deposition [14], oxidative stress reduction [15], brain perfusion increase, and upregulation of neurotrophic factors [16] are a few of the widely studied and reported effects. These effects were presented as improved mood state, improved cognitive function, reduced comorbidities, and decreased risk of falls [17,18].

Combined physical and cognitive training may facilitate the neuroplasticity potential and enhance an individual's capacity to respond to new demands, resulting in mutual enhancement [19].

Recent advances in information and communication technologies (ICTs) and health informatics offer new and elderly-friendly training on web-based platforms [20]. These platforms may also serve people with limited access to an organized day care center, in their own home or an assisted living/nursing home facility, occasionally from a distance with remote surveillance by specialized personnel. Technology-assisted solutions for elderly physical training through gaming, termed exergaming, have been increasingly investigated [21,22]. Validation of the effectiveness of these approaches is currently a top research priority [23].

Apolipoprotein E

The ε4 allele of the apolipoprotein E (APOE) gene (APOE ε4) is the major genetic risk factor for Alzheimer disease. APOE ε4 carriers not only have a higher risk but also an earlier onset of Alzheimer disease by 10 to 20 years [24] in a gene dose-dependent manner. APOE ε4 (ε4/ε4) homozygotes compared with persons homozygous for risk-neutral APOE ε3 (ε3/ε3) may have up to 15 times the increased the risk for developing Alzheimer disease while APOE ε4 heterozygotes (ε4/ε3 or ε4/ε2) only have a 4 times higher risk [25]. Numerous studies have attempted to elucidate the underlying mechanism for APOE ε4 influences on Alzheimer disease onset and progression. It has been difficult to determine whether the APOE ε4 represents a gain of toxic function, a loss of neuroprotective function, or both [26]. It is noteworthy that APOE ε4 is associated not only with Alzheimer disease but also with altered brain metabolism and structure in young cognitively normal adults [27].

The presence of APOE ε4 significantly influences the progression of healthy elderly to MCI and Alzheimer disease, and the progression risk peaks between ages 70 and 75 years [28]. Studies are suggesting that the effect of the ε4 allele on cognitive decline is stronger in this earlier clinical stage in comparison with later and more severe stages [26]. Moreover, the more sedentary the lifestyle of the elderly person is, the higher the impact of APOE ε4 on cerebral amyloid deposition. However, not all APOE ε4 carriers will develop Alzheimer disease, thereby suggesting the interactive effects of APOE genotype with other genetic or environmental factors [25].

Our long-term study aims to investigate the cognitive effects of different computer-based interventions depending on the APOE isoform. Also, the follow-up reevaluations at 6, 12, and 24 months intend to determine which training program, if any, can postpone further cognitive decline and dementia onset. In this paper, which describes the first part of the study, the pre-post training evaluation of the interventional groups is presented. We assessed and compared the efficacy of the different web-based interventions and subsequently assessed whether the APOE genotype may influence the outcome. It is

expected that such a finding may be useful for the improvement of the currently existing and future designs of web-based, technology-assisted therapeutic interventions.

Methods

Study Design

We used a quasi-experimental study design [29] to explore the efficacy of different interventions on participants with MCI. Participants were allocated between 2009 and 2017 into three different interventional groups. Based on previous efficacy evidence of the Long-Lasting Memories (LLM) intervention, an integrated ICT platform combining cognitive exercises with physical activity [19,30], we did not use a passive group in the pre-post assessment due to ethical considerations. Thirteen participants, however, were evaluated in the same pre- and postintervention period without participating in the intervention for personal reasons. All other passive data were retrospectively collected from the database records of the Greek Association of Alzheimer's Disease and Related Disorders, matched for demographic data and initial diagnosis. This passive group was used only for the long-term follow-up.

Randomization of the participants was not feasible mainly for practical issues, time, and financial limitations of the study. Allocation to groups was driven by nonsystematic practical and logistic reasons (national holiday time, number of successfully screened participants at a given time point, etc) but was not influenced by participant choice, motivation, or compliance. The type of intervention applied each time in each place (spiritual center, open care center) was determined before the initiation of the screening procedure, and it was an open call to the elderly whether they were cognitively intact or not. Interventions took place both in the east and west areas of Thessaloniki, minimizing geographic socioeconomic differences of participants. From all the participants who enrolled in the LLM project, we analyzed those with an initial diagnosis of MCI fulfilling the selection criteria.

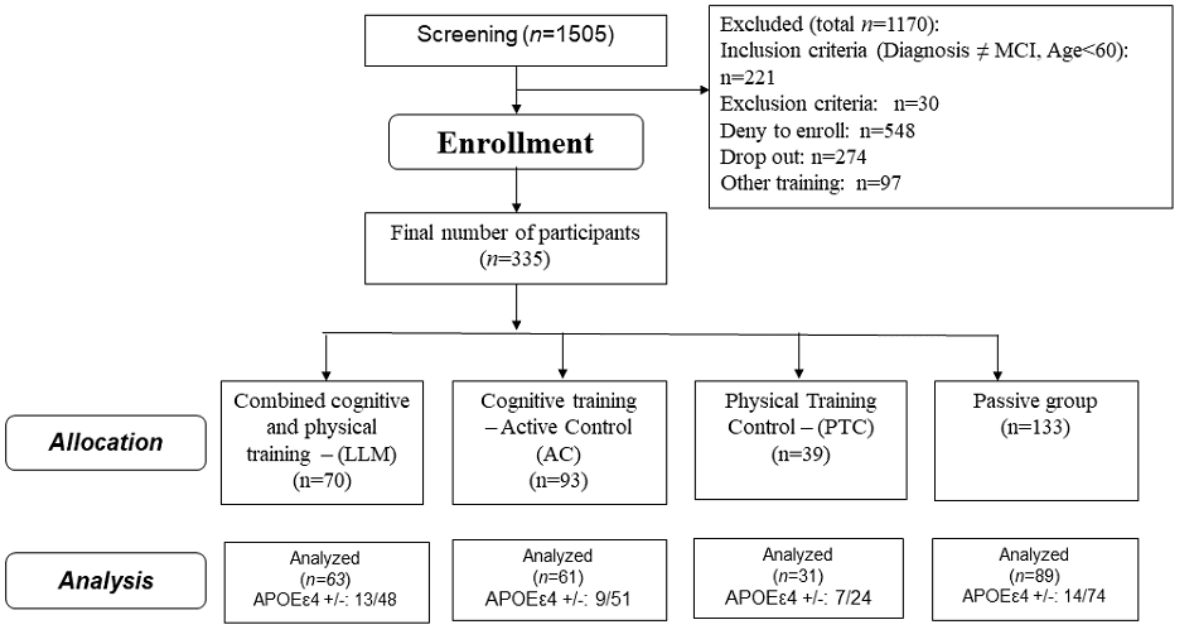
The training lasted about 8 to 12 weeks, and participants completed at least 24 sessions of cognitive training and 16 sessions of physical training. Screening evaluations were conducted 1 to 2 weeks before interventions, while posttest evaluations occurred 1 to 2 weeks after the end of the training. Neuropsychologists performing the pre- and postintervention neuropsychological assessments were generally different from those who administered the program.

Participant performance was assessed and compared among the different interventional groups to investigate a potential superiority of an intervention. Moreover, the performance of each interventional group, in each test between the two time points (pre-post), was assessed separately, investigating specific cognitive domain improvement. A second analysis, based on genotype, was performed subsequently. Each interventional group was subdivided into APOE ε4 carriers and non–APOE ε4 carriers. A comparison of each APOE subgroup among the interventional groups was performed to assess potential differences in efficacy in different genotypes. The performance of each APOE subgroup in the two time points was assessed separately within the interventional group as well.

Participants

A total of 215 MCI participants (intervention 202 [LLM 70; Active Control, or AC 93; Physical Training Control, or PTC 39] and passive 13; single and multiple domains) were recruited during a thorough screening procedure. The rest of the passive group data (n=120) was collected retrospectively as described in Study Design section (Figure 1). Males and females aged 60 years and older, fluent in Greek, were invited to participate. The call was made in church spiritual centers, open care centers for the elderly in east and west areas of Thessaloniki, and day care centers of the Greek Association of Alzheimer's Disease and Related Disorders. Participants provided written informed consent and were compensated for their participation in the study. The study protocol was approved by the bioethics committee of the School of Medicine of the Aristotle University of Thessaloniki (protocol no 38/5.6.2013).

Figure 1. Study flowchart with the number of the participants screened and allocated in each interventional group. MCI: mild cognitive impairment; LLM: Long-Lasting Memories; APOE: apolipoprotein E.



Exclusion criteria included any severe physical illness, current psychiatric or other neurological disorder (stroke, multiple sclerosis, Parkinson's disease, epilepsy, traumatic brain injury, etc), history of drug or alcohol abuse, and use of neuromodifying drugs (other than cholinesterase inhibitor).

All participants reported normal or corrected-to-normal vision and hearing. Before the training, all participants underwent an extensive neuropsychological evaluation performed using standardized Greek versions: Mini-Mental State Examination (MMSE) [31,32], Montreal Cognitive Assessment (MoCA) [33,34], and Trail Making Test, Part B (Trail B) [35,36] to examine task-switching, processing speed, and visuospatial ability; California Verbal Learning Test (CVLT) [37-39] and Rey Auditory Verbal Learning Test (RAVLT) in order to measure the ability of learning, long-term memory, and verbal episodic memory [40,41]; digit span forward and backward [42,43] to assess working memory; Instrumental Activities of Daily Living scale (IADL) [44,45], Functional Rating Scale of Symptoms of Dementia (FRSSD) [46], and Functional and Cognitive Assessment Test (FUCAS) [47] to assess daily functionality; Beck Depression Inventory (BDI) [48,49] and Geriatric Depression Scale (GDS) [50,51] to measure depression; and Beck Anxiety Inventory (BAI) [52,53] for anxiety evaluation. Participants were also subjected to blood tests and brain magnetic resonance imaging. A neurologist evaluated the neuropsychological, medical, and laboratory results. The Cumulative Illness Rating Scale-Geriatric (CIRS) was calculated based on their medical history to assess participant comorbidity [54]. Diagnosis and categorization were based on clinical criteria [1,55] and were made by a dementia expert neurologist (MT). A detailed description of the procedure is described by Bamidis et al [19].

Interventions

The first group attended the combined cognitive and physical training via the LLM project, using an integrated web service system through a universal interface, facilitated by touch screen systems [23,56]. Cognitive training was performed using a Greek-adapted version of the Brain Fitness software (later branded as BrainHQ, Posit Science Corporation) [57], and all physical exercises were implemented in the FitForAll (FFA) platform, an innovative ICT exergaming platform designed by Aristotle University of Thessaloniki [23]. FFA has four levels of difficulty and combines aerobic exercise; exercises of endurance, strength, and balance; and a cooling down process to recover normal cardiac rhythm.

The second group was the AC group, which was exposed to a homemade computerized cognitive training software suite (Video GRade, Lab of Medical Physics at the Aristotle University of Thessaloniki [58]), encompassing short documentary videos with themes from nature, art, history, and culture. At the end of each video, participants digitally performed a multiple-choice questionnaire about the documentary, following the same training dosage as the first group. The third group was the PTC group, which was exposed to computerized physical training using the FFA platform for the same number of sessions as the others. The fourth group was the passive control group, which did not follow any training program.

Apolipoprotein E Genotyping

Blood samples used for genotyping were collected in EDTA-containing receptacles. DNA was extracted from peripheral blood using the QIAamp Blood DNA purification kit (Qiagen Inc). To determine the APOE genotype, part of the APOE gene (228 bp) containing both polymorphic sites (amino

acid positions 112 and 158) was amplified by polymerase chain reaction analysis using the following primers: forward: 5″-GGCACGGCTGTCCAAGGAGCTGCA-3″ and reverse: 5″-GCCCCGGCCTGGTACACTGCCAG-3″, according to the method described by Koutroumani et al [59].

Statistical Analysis

Multiple Imputation

In clinical and epidemiological research, the problem of missing data is almost unavoidable. In our study, missing data were handled using multiple imputation [60] tackling the missing data problem from three aspects: (a) the missing data proportion, (b) the mechanisms of incomplete data, and (c) the missing data patterns as suggested by Dong and Peng [61]. The missing data proportion was calculated, and missing data mechanisms were assessed using the Little multivariate test [62] and performing *t* tests of mean differences between the complete and missing data groups following the guidelines illustrated with a sample dataset from IBM SPSS Missing Values 20 [63]. The missing data pattern was explored using the command Analyze Patterns, which provides descriptive measures of the missing data patterns and could be useful as an exploratory step (see Multimedia Appendix 1 for methodology details). The analysis was performed using SPSS Statistics version 23 (IBM Corporation).

In our model, demographic data along with the scores of different neuropsychological tools and questionnaires were included following the guidelines incorporated in Dong and Peng [61]. Twenty imputations were chosen to remove noise from estimations, such as reducing sampling variability from the imputation procedure [64]. Following the completion of the multiple imputation, the 20 complete datasets were handled by standard statistical procedures after splitting the imputed dataset based on the imputation number. As many procedures in SPSS Statistics do not support pooling, we decided to report mainly pooled mean ranks that are naïve pooled. Even though we lose some of the descriptive power we could get from medians, we still gain valuable information about our groups in terms of our dependent variables. Moreover, we calculated the averaged median across the 20 complete datasets, as suggested by van Ginkel and Kroonenberg [65].

Demographics

During the baseline neuropsychological evaluation, a battery of tests was administered to participants including various neuropsychological tests and questionnaires assessing different cognitive aspects, performance in daily life activities, and the affective state of the participants. The scores collected by the tests and questionnaires used were tested for the normality assumption similarly to demographic data. As scores were not approximately normally distributed in all imputations, nonparametric analysis (Kruskal-Wallis test) was employed to explore differences between groups. When differences among groups reached statistical significance ($P<.05$), the Mann-Whitney *U* test was performed pairwise. Alpha inflation due to multiple comparisons was prevented using a Bonferroni correction.

Performance of participants was further discriminated depending on the presence of the APOE ε4 genotype. More precisely, scores on neuropsychological tests and questionnaires were also compared between APOE ε4 carriers versus non–APOE ε4 carriers. The normality assumption of test scores was explored, performing the described methodology between the two groups. As normality assumption was not met for both groups, a nonparametric analysis was performed (Mann-Whitney *U* test).

Neuropsychological Evaluation

Among-Group Analysis

Neuropsychological evaluations were administrated to the intervention groups (LLM, AC, and PTC) before and after the training completion. Scores of neuropsychological tests and questionnaires were analyzed using the group (LLM, AC, and PTC) as between factor and the time (pre- and posttraining) as within factor. The assumptions of mixed-model analysis of variance were not met in each cell of the design. Thus, an alternative analysis was followed using nonparametric procedures. Differences in scores at the two time points were computed for each test and questionnaire and then tested for normality. The Kruskal-Wallis H test was performed in differences (post-pre) of scores between groups. When differences among groups reached statistical significance ($P<.05$), the Mann-Whitney *U* test was performed pairwise. In this case, *P* values were corrected for multiple comparisons using a Bonferroni correction.

Within-Group Analysis

After grouping our dataset by the imputation number and group, Wilcoxon signed-rank tests were run to investigate possible differences of each group at different time points (pre- and posttraining) in M (M=20) complete datasets. Additionally, we examined a small sample of the passive group (n=13) regarding their performance in several tests such as the MMSE, MoCA, CVLT total, Trail B, IADL, and GDS at the two time points. Depending on normality assumption, different analyses were performed: either paired *t* test or Wilcoxon signed-rank test.

Among-Group Analysis Based on Genotype

Data were split not only by imputation number but also by genotype (carriers, non-ε4 carriers). Differences in scores at the two time points were computed for each test and questionnaire and then tested for normality. As normality assumption was not fulfilled for all imputations and groups, respectively, nonparametric analysis for between-group comparison was determined. Kruskal-Wallis H tests were run for the post-pre differences in scores having as grouping variable the group (LLM, AC, PTC). Results were split depending on the genotype. Significant differences among groups were explored running post hoc tests (pairwise comparisons using Mann Whitney *U* tests) and correcting for multiple comparisons.

Within-Group Analysis Based on Genotype

Data were split by imputation number, genotype (APOE ε4 carriers, non-ε4 carriers), and group. As violations of normality were observed, nonparametric analysis for within-comparison (pre, post) was performed.

Results

Demographics

A total of 335 participants in total were allocated to one of the four different groups (LLM, AC, PTC, passive). The baseline comparisons revealed significant differences between the four groups regarding demographics and cognitive performance. Therefore, a smaller sample of 244 subjects without differences in baseline evaluation scores was used for the pre-post comparison to evaluate the potential effect of the intervention without major confounding factors. Between and within-group comparisons at the first analysis stage were performed in LLM, AC, and PTC participants (see Table 1).

Table 1. Demographic characteristics of participants included in the study.

Groups	LLM[a] n=70	AC[b] n=93	PTC[c] n=39	Passive n=133	Test results	
					χ^2_3	*P* value
Analyzed n=244	63	61	31	89		
Age in years					6.8	.08
Median	69	69	67	68		
IQR	7	9	8	14		
Q1, Q3	66.0, 73.0	65.0, 74.0	63.0, 71.0	60.0, 74.0		
Gender, n					5.7	.13
Male	51	46	28	63		
Female	12	15	3	26		
Education in years					7.5	.06
Median	6	8	6	8		
IQR	7	6	1	6		
Q1, Q3	6.0, 13.0	6.0, 12.0	6.0, 7.0	6.0, 12.0		
Comorbidity index (CIRS[d]), pooled mean ranks	121.39	105.11	129.53	132.76	6.4	.11
APOE[e] ε4/- (frequency % in total sample of 335 participants)						
APOE ε4 carriers	15 (21.4)	13 (14.0)	9 (23.1)	30 (22.6)		
Non-APOE ε4 carriers	53 (75.7)	77 (82.8)	30 (76.9)	101 (75.9)		
APOE ε4/- (frequency % in total sample of 244 participants)						
APOE ε4 carriers	13 (20.6)	9 (14.8)	7 (22.6)	14 (15.7)		
Non-APOE ε4 carriers	48 (76.2)	51 (83.6)	24 (77.4)	74 (83.1)		

[a]LLM: Long-Lasting Memories.

[b]AC: Active Control.

[c]PTC: Physical Training Control.

[d]CIRS: Cumulative Illness Rating Scale–Geriatric.

[e]APOE: apolipoprotein E.

Discriminating the participants with regard to the presence of APOE ε4 genotype, we found that there were 43 APOE ε4 carriers, 197 non–APOE ε4 carriers, and in 4 cases there were missing data. Planned analysis of available data showed that the two groups did not differ in age (*U*=4022.00, *P*=.60), education years (*U*=4041.50, *P*=.62), and CIRS scores (*U*=3267.00, *P*=.69). Additionally, the two independent binomial proportions regarding the proportion of gender across groups were statistically significantly different (χ^2_1=3.9, *P*=.048; see Table 1).

In more detail, participants who were APOE ε4 carriers were barely older (age [pooled mean ranks] APOE ε4 carriers: 125.47; non–APOE ε4 carriers: 119.42) and more educated than noncarriers (education years [pooled mean ranks] APOE ε4 carriers: 125.01; non–APOE ε4 carriers: 119.52). Moreover, they had elevated CIRS scores relative to noncarriers (CIRS [pooled mean ranks] APOE ε4 carriers: 123.83; non–APOE ε4 carriers: 119.77).

Neuropsychological Evaluation

Among-Group Analysis

Nonsignificant changes have been observed between groups in most of the neuropsychological tests comparing the scores' differences at the two time points. Significant differences have been found only in the post-pre comparisons on the MMSE and GDS tests (Table 2). Based on the post hoc analysis, the LLM group seems to have a significantly lower performance on

MMSE compared with AC participants and considerably more depressive symptoms relative to the PTC group.

Table 2. Presentation of among-group comparison results along with descriptive measures (pooled mean ranks of post-pre scores) for each group (Long-Lasting Memories, Active Control, and Physical Training Control).

Neuropsychological test	LLM[a]	AC[b]	PTC[c]	Test results	
	Pooled mean ranks	Pooled mean ranks	Pooled mean ranks	χ^2_2	*P* value
Cognitive domain					
MMSE[d]	69.47	89.58	72.54	7.1	.04
MoCA[e]	71.88	81.26	84.03	2.5	.34
RAVLT1[f]	77.20	78.81	78.04	1.6	.52
RAVLT total[g]	74.84	79.24	82.00	1.9	.48
RAVLTD[h]	77.97	79.08	75.94	1.5	.53
CVLT1[i]	69.37	84.09	83.57	5.0	.14
CVLT total[j]	71.58	84.77	77.71	3.4	.26
CVLTD[k]	71.33	79.23	89.14	4.7	.20
Trail B[l]	74.95	81.23	77.84	1.0	.67
Digit span forward	75.15	77.65	84.48	1.5	.52
Digit span backward	72.73	78.77	87.19	2.6	.32
Functionality					
FUCAS[m]	72.68	81.63	81.67	2.4	.39
FRSSD[n]	74.47	83.83	73.70	3.9	.30
IADL[o]	77.19	80.15	75.41	1.3	.61
Affective domain					
GDS[p]	88.33	78.50	56.02	8.5	.005
BAI[q]	76.33	80.74	75.99	0.1	.51
BDI[r]	79.18	79.41	72.83	4.8	.49

[a]LLM: Long-Lasting Memories.
[b]AC: Active Control.
[c]PTC: Physical Training Control.
[d]MMSE: Mini-Mental State Examination.
[e]MoCA: Montreal Cognitive Assessment.
[f]RAVLT1: Rey Auditory Verbal Learning Test: immediate recall.
[g]RAVLT total: Rey Auditory Verbal Learning Test: sum up of 5 recall attempts.
[h]RAVLTD: Rey Auditory Verbal Learning Test: delayed recall.
[i]CVLT1: California Verbal Learning Test: immediate recall.
[j]CVLT total: California Verbal Learning Test: sum up of 5 recall attempts.
[k]CVLTD: California Verbal Learning Test: delayed recall.
[l]Trail B: Trail Making Test, Part B.
[m]FUCAS: Functional and Cognitive Assessment Test.
[n]FRSSD: Functional Rating Scale of Symptoms of Dementia.
[o]IADL: Instrumental Activities of Daily Living scale.
[p]GDS: Geriatric Depression Scale.
[q]BAI: Beck Anxiety Inventory.
[r]BDI: Beck Depression Inventory.

Within-Group Analysis

Long-Lasting Memories Group

LLM participants scored significantly higher in RAVLT1, RAVLT total, CVLT1, CVLT total, and CVLTD tests after their training compared with the baseline (Table 3). The within-group comparison did not reveal any significant differences either in functionality test scores of FUCAS, FRSSD, and IADL tests or emotion test scores of GDS, BAI, and BDI scores when comparing pre- and posttraining scores (Multimedia Appendix 2 Table A).

Active Control Group

The AC group showed significant improvement in their performance on the MMSE, RAVLT total, CVLT1, CVLT total, CVLTD, and digit span backward test when comparing the test scores before and after their training. They also showed significantly higher scores in the FUCAS test after training compared with the baseline evaluation, possibly indicating a decrease in their functionality in daily life activities (Table 3). Significant changes in the performance of the AC group at GDS, BAI, and BDI tests were not found when comparing scores at the two time points (Multimedia Appendix 2 Table B).

Physical Training Control Group

The PTC participants scored significantly higher in MoCA, RAVLT total, CVLT total, CVLTD, digit span forward, and digit span backward tests at the posttraining screening relative to the baseline. A significant decrease was also observed in scores on the GDS test at the posttraining neuropsychological screening relative to those of the baseline evaluation (Table 3; see Multimedia Appendix 3 Figure i for detailed results). Functionality scores of FUCAS, FRSSD, and IADL tests did not change significantly at the two time points (Multimedia Appendix 2 Table C).

The passive group (n=13), which was reassessed after the 12 weeks, did not reveal significant changes in their performance on the tests MMSE (t_{12}=2.082; *P*=.059), MoCA (W=–0.319; *P*=.75), Trail B (t_{11}=–0.656; *P*=.53), IADL (W=–0.577; *P*=.56), and GDS (W=–0.852; *P*=.39) at the two time points. However, a significant increase of 8.385 score units (95% CI 3.39 to 13.38, Cohen *d*=1.015 [38]) was found in CVLT total (t_{12}=3.659; *P*=.003; CVLT baseline: 38.00; CVLT after 12 week: 46.38).

Table 3. Test scores with significant improvement for each interventional group when comparing their scores at the two time points.

Neuropsychological test	Pooled mean ranks: negative[a]	Pooled mean ranks: positive[b]	Averaged median before training	Averaged median after training	Test results	
					Wilcoxon signed-rank test	*P* value
LLM[c]						
RAVLT1[d]	23.88	34.41	4.535	5.428	–3.2	.045
RAVLT total[e]	25.45	32.65	37.121	41.85	–3.2	.04
CVLT1[f]	20.35	29.62	4.795	5.311	–2.4	.04
CVLT total[g]	25.22	31.18	41.369	46.206	–3.4	.002
CVLTD[h]	19.67	30.73	8.559	9.239	–2.7	.02
AC[i]						
MMSE[j]	22.56	24.96	27.000	28.000	–4.1	<.001
RAVLT total	24.36	33.47	38.903	45.507	–3.8	.005
CVLT1	20.36	30.79	4.584	6.650	–4.0	<.001
CVLT total	16.94	32.25	39.941	47.742	–4.6	<.001
CVLTD	23.74	30.98	8.259	9.908	–3.6	.002
Digit span backward	21.30	26.96	4.004	4.488	–2.5	.03
PTC[k]						
MoCA[l]	9.99	13.73	22.394	23.000	–2.0	.04
RAVLT total	12.43	17.51	37.546	44.117	–2.7	.02
CVLT total	11.56	16.41	41.187	48.359	–2.6	.02
CVLTD	12.32	16.08	8.175	10.869	–3.3	.02
Digit span forward	10.66	14.02	5.000	5.000	–2.3	.02
Digit span backward	9.12	12.16	4.000	4.000	–2.8	.006
GDS[m]	13.46	7.58	2.000	0	–3.5	.001

[a]Negative mean rank: test score post < test score pre.

[b]Positive mean rank: test score post > test score pre.

[c]LLM: Long-Lasting Memories.

[d]RAVLT1: Rey Auditory Verbal Learning Test: immediate recall.

[e]RAVLT total: Rey Auditory Verbal Learning Test: sum up of 5 recall attempts.

[f]CVLT1: California Verbal Learning Test: immediate recall.

[g]CVLT total: California Verbal Learning Test: sum up of 5 recall attempts.

[h]CVLTD: California Verbal Learning Test: delayed recall.

[i]AC: Active Control.

[j]MMSE: Mini-Mental State Examination.

[k]PTC: Physical Training Control.

[l]MoCA: Montreal Cognitive Assessment.

[m]GDS: Geriatric Depression Scale.

Among-Group Analysis Based on Genotype

Non-ε4 Carriers Among Groups

Non-ε4 carriers appear to significantly alter their performance neither in any cognitive tests nor in any test assessing their functionality in activities of daily living at the two time points depending on the group. Significant among-group differences were observed only in GDS tests. Planned post hoc tests revealed that LLM showed a considerably greater change in their geriatric depressive symptoms than the PTC group (Table 4 and Multimedia Appendix 3 Figure i).

Apolipoprotein E ε4 Carriers Among Groups

The ε4 carriers did not change their cognitive status significantly depending on the training given as evaluated by post-pre differences between groups. Furthermore, ε4 carriers seem to

preserve their functionality and mood status as evaluated by respective tests at the two time points (Table 5 and Multimedia Appendix 3 Figure ii).

Table 4. Among-group comparison results in non-ε4 carriers. Descriptive measures (pooled mean ranks of post-pre scores) for each group (Long-Lasting Memories, Active Control, and Physical Training Control) are displayed.

Neuropsychological tests	LLM[a]	AC[b]	PTC[c]	Test results	
	Pooled mean ranks	Pooled mean ranks	Pooled mean ranks	χ^2_2	*P* value
Cognitive domain					
MMSE[d]	57.11	70.00	54.77	4.6	.12
MoCA[e]	60.75	64.68	56.39	0.5	.80
RAVLT1[f]	61.13	62.48	62.73	1.7	.49
RAVLT total[g]	59.67	62.69	65.19	1.8	.51
RAVLTD[h]	62.93	61.79	60.60	1.7	.54
CVLT1[i]	53.86	67.73	66.11	5.3	.12
CVLT total[j]	56.74	67.69	60.43	3.2	.28
CVLTD[k]	57.78	61.58	71.33	3.6	.26
Trail B[l]	61.96	64.68	56.39	1.3	.58
Digit span forward	58.62	61.85	69.07	1.8	.44
Digit span backward	57.62	63.21	68.18	1.8	.45
Functionality					
FUCAS[m]	55.59	65.65	67.07	3.3	.27
FRSSD[n]	58.73	66.00	60.05	3.1	.35
IADL[o]	59.63	65.26	59.80	2.1	.48
Affective domain					
GDS[p]	69.64	61.65	47.46	6.6	.047
BAI[q]	60.18	64.09	56.13	1.9	.55
BDI[r]	64.13	62.76	56.13	2.3	.46

[a]LLM: Long-Lasting Memories.
[b]AC: Active Control.
[c]PTC: Physical Training Control.
[d]MMSE: Mini-Mental State Examination.
[e]MoCA: Montreal Cognitive Assessment.
[f]RAVLT1: Rey Auditory Verbal Learning Test: immediate recall.
[g]RAVLT total: Rey Auditory Verbal Learning Test: sum up of 5 recall attempts.
[h]RAVLTD: Rey Auditory Verbal Learning Test: delayed recall.
[i]CVLT1: California Verbal Learning Test: immediate recall.
[j]CVLT total: California Verbal Learning Test: sum up of 5 recall attempts.
[k]CVLTD: California Verbal Learning Test: delayed recall.
[l]Trail B: Trail Making Test, Part B.
[m]FUCAS: Functional and Cognitive Assessment Test.
[n]FRSSD: Functional Rating Scale of Symptoms of Dementia.
[o]IADL: Instrumental Activities of Daily Living scale.
[p]GDS: Geriatric Depression Scale.
[q]BAI: Beck Anxiety Inventory.
[r]BDI: Beck Depression Inventory.

Table 5. Among-group comparison results in ε4 carriers. Descriptive measures (pooled mean ranks of post-pre scores) for each group (Long-Lasting Memories, Active Control, and Physical Training Control) are presented.

Neuropsychological tests	LLM[a]	AC[b]	PTC[c]	Test results	
	Pooled mean ranks	Pooled mean ranks	Pooled mean ranks	χ^2_2	*P* value
Cognitive domain					
MMSE[d]	12.44	17.63	16.37	2.3	.31
MoCA[e]	12.25	16.20	18.57	3.2	.25
RAVLT1[f]	14.99	15.52	14.36	1.1	.66
RAVLT total[g]	14.57	14.94	15.88	1.8	.49
RAVLTD[h]	14.48	15.96	14.74	1.6	.57
CVLT1[i]	14.09	14.67	17.12	1.6	.49
CVLT total[j]	13.30	16.02	16.84	1.7	.50
CVLTD[k]	13.37	15.69	17.15	1.9	.48
Trail B[l]	12.16	15.52	19.61	4.1	.18
Digit span forward	15.50	13.51	15.99	1.6	.61
Digit span backward	14.42	14.68	16.49	0.8	.72
Functionality					
FUCAS[m]	15.21	15.67	13.74	1.2	.59
FRSSD[n]	15.25	15.99	13.25	2.0	.53
IADL[o]	16.62	12.58	15.10	2.9	.45
Affective domain					
GDS[p]	16.87	17.29	8.57	5.7	.07
BAI[q]	15.07	15.86	13.77	1.8	.56
BDI[r]	14.68	15.48	14.98	1.0	.64

[a]LLM: Long-Lasting Memories.
[b]AC: Active Control.
[c]PTC: Physical Training Control.
[d]MMSE: Mini-Mental State Examination.
[e]MoCA: Montreal Cognitive Assessment.
[f]RAVLT1: Rey Auditory Verbal Learning Test: immediate recall.
[g]RAVLT total: Rey Auditory Verbal Learning Test: sum up of 5 recall attempts.
[h]RAVLTD: Rey Auditory Verbal Learning Test: delayed recall.
[i]CVLT1: California Verbal Learning Test: immediate recall.
[j]CVLT total: California Verbal Learning Test: sum up of 5 recall attempts.
[k]CVLTD: California Verbal Learning Test: delayed recall.
[l]Trail B: Trail Making Test, Part B.
[m]FUCAS: Functional and Cognitive Assessment Test.
[n]FRSSD: Functional Rating Scale of Symptoms of Dementia.
[o]IADL: Instrumental Activities of Daily Living scale.
[p]GDS: Geriatric Depression Scale.
[q]BAI: Beck Anxiety Inventory.
[r]BDI: Beck Depression Inventory.

Within-Group Analysis Based on Genotype

Long-Lasting Memories Group

The ε4 carriers in the LLM group significantly improved their performance at Trail B while non-ε4 carriers considerably improved their performance on many cognitive tests such as the MMSE, MoCA, RAVLT1, RAVLT total, CVLT1, CVLT total, and CVLTK, comparing their test scores both before and after the training (Table 6 and Multimedia Appendix 3 Figure iii). Neither ε4 carriers nor non-ε4 carriers considerably altered their functional status in activities of daily living and depressive and anxiety symptomatology (Multimedia Appendix 2 Table D).

Active Control Group

AC ε4 carriers did not significantly change their scores on cognitive tests before and after training (Multimedia Appendix 2 Table E), while non-ε4 carriers showed a considerable improvement on the MMSE, MoCA, digit span backward, RAVLT total, an on different categories of the CVLT test. The ε4 carriers did not considerably change their functionality in activities of daily living as assessed by different tests at the two time points while the non-ε4 carriers significantly altered their scores on the FUCAS test, indicating a decrease in their functionality (Table 6 and Multimedia Appendix 3 Figure iv). Nonsignificant changes were observed in both ε4 carriers and non-ε4 carriers with regard to their depressive and anxiety symptomatology (Multimedia Appendix 2 Table E).

Physical Training Control

The ε4 carriers of PTC scored significantly better on digit span backward tests after training compared with the baseline, while non-ε4 carriers of the same group showed considerable improvement in a couple of tests such as RAVLT total, CVLT total, CVLTK, digit span forward, and digit span backward (Table 6 and Multimedia Appendix 3 Figure v).

PTC ε4 carriers did not show a significant change in their functionality in activities of daily living as assessed by the FUCAS, FRSSD, and IADL tests (Multimedia Appendix 2 Table F). However, non-ε4 carriers showed a marginally significant change only in FUCAS scores when comparing their scores at the two time points. Both ε4 carriers and non-ε4 carriers seemed to significantly improve their geriatric depressive scores after the training compared with the baseline screening (Table 6). Scores on BAI and BDI tests at the two time points did not change significantly in either group (Multimedia Appendix 2 Table F).

Table 6. Significant score changes in the performance of ε4 carriers and non-ε4 carriers of the Long-Lasting Memories, Active Control, and Physical Training Control groups.

Neuropsychological tests	Pooled mean ranks: negative[a]	Pooled mean ranks: positive[b]	Averaged median before training	Averaged median after training	Test results	
					Wilcoxon signed-rank test	*P* value
LLM[c] ε4 carriers						
Trail B[d]	7.03	6.87	234.797	185.516	–2.2	.03
LLM non-ε4 carriers						
MMSE[e]	14.30	20.36	27.000	28.000	–2.0	.047
MoCA[f]	20.24	22.85	22.915	24.000	–2.2	.04
RAVLT1[g]	17.33	26.15	4.523	5.463	–3.0	.049
RAVLT total[h]	18.95	26.06	37.699	43.220	–3.0	.04
CVLT1[i]	14.47	21.42	4.751	5.566	–2.6	.03
CVLT total[j]	18.88	22.97	41.605	46.994	–3.4	.002
CVLTD[k]	14.49	22.83	8.471	9.598	–2.7	.02
Active non-ε4 carriers						
MMSE	19.58	21.34	26.948	28.000	–3.8	<.001
MoCA	17.28	27.77	22.997	24.726	–2.4	.04
RAVLT total	19.28	28.41	38.771	45.806	–3.6	.01
CVLT1	14.54	26.21	4.507	6.832	–4.2	<.001
CVLT total	12.45	26.75	39.180	48.408	–4.5	<.001
CVLTD	19.06	25.64	8.312	10.002	–3.5	.002
Digit span backward	17.41	22.16	4.000	4.362	–2.3	.046
FUCAS[l]	17.66	23.69	43.798	44.271	–3.2	.004
PTC[m] ε4 carriers						
Digit span backward	2.50	3.75	4.000	5.000	–2.1	.03
GDS[n]	3.50	0	2.000	0	–2.2	.03
PTC non-ε4 carriers						
RAVLT total	9.37	13.62	37.509	44.118	–2.6	.03
CVLT total	9.32	13.02	41.230	47.803	–2.4	.04
CVLTD	9.41	13.16	7.871	10.822	–3.0	.03
Digit span forward	9.02	10.96	4.405	5.000	–2.5	.01
Digit span backward	6.52	9.96	4.000	4.000	–2.2	.03
FUCAS	8.54	14.55	43.537	44.505	–2.3	.05
GDS	10.68	5.28	2.000	0	–2.8	.006

[a]Negative mean rank: test score post < test score pre.

[b]Positive mean rank: test score post > test score pre.

[c]LLM: Long-Lasting Memories.

[d]Trail B: Trail Making Test, Part B.

[e]MMSE: Mini-Mental State Examination.

[f]MoCA: Montreal Cognitive Assessment.

[g]RAVLT1: Rey Auditory Verbal Learning Test: immediate recall.

[h]RAVLT total: Rey Auditory Verbal Learning Test: sum up of 5 recall attempts.

[i]CVLT1: California Verbal Learning Test: immediate recall.

[j]CVLT total: California Verbal Learning Test: sum up of 5 recall attempts.

[k]CVLTD: California Verbal Learning Test: delayed recall.

[l]FUCAS: Functional and Cognitive Assessment Test.

[m]PTC: Physical Training Control.

[n]GDS: Geriatric Depression Scale.

Discussion

Principal Findings

To our knowledge, this is the first quasi-experimental study to investigate in elders with MCI the impact of a combination of computerized physical and cognitive training not only in terms of cognitive decline in general but also based on the different APOE isoforms. Our results indicate that MCI APOE ε4 carriers respond differently and less prominently in web-based interventions.

Since there are no previous data about nonpharmacological computerized interventional outcomes on APOE ε4 carriers, we cannot possibly make direct comparisons with past research, but we do underline the importance of these new findings and the likely key role of future investigations with longitudinal randomized trials.

There are, however, recent data on computerized training interventions in MCI subjects that present results in line with our findings, suggesting improvement of learning ability (RAVLT total, CVLT total) and short-term memory (RAVLT1, CVLT1, digit span), verbal memory (RAVLT, CVLT), task-switching, processing speed and visuospatial memory (Trail B), episodic memory (RAVLTD and CVLTD delayed recall) and attention (MMSE, MoCA), and a positive effect on depressive symptoms (GDS) as well (for a review see Klimova et al [66]). Although these outcomes appear promising, researchers underline the limitations of these studies because of their duration, small sample sizes, and methodological differences. A recent meta-analysis on classic cognitive interventions targeting multiple domains in MCI subjects presented cognitive benefits in working memory, attention, and verbal memory in a larger study series [5]. The above outcomes, which are consistent with ours, enhance the strength of the thesis that training in the MCI population is a promising tool against neurodegeneration.

The combination of physical and cognitive training has also been studied previously in this high-risk population, underlining not only the cognitive benefits [67,68] but also the improvement of biological parameters such as brain-derived neurotrophic factor levels, grey matter volume [69], and increased para-hippocampal cerebral blood flow [67].

Recent data on exergaming, using the latest technology of the virtual reality with low and high cognitive engagement, also demonstrated improvement in verbal memory, a cognitive parameter that showed improvement in all of our interventional groups [69].

In our study, the analysis among the groups did not reveal statistically significant differences in the effect of the three different interventions other than the MMSE (LLM<AC) and the GDS (LLM>PTC), so we could not possibly identify the best intervention. Assessment of activities of daily living revealed no differences, while changes in the affective domain underlined the positive effect of physical exercise on stress and depression.

Within each interventional group, statistically significant differences were more obvious. The LLM group showed significant improvement in episodic memory, learning ability, and long-term memory, while no significant changes existed in functionality and the affective domain. The AC group showed a considerable improvement in global cognition status as assessed by MMSE, episodic memory, learning ability, long-term memory, and working memory. Finally, the PTC group showed significant improvement in global cognitive status as assessed by MoCA and in working memory, learning ability, and long-term memory. The PTC group also had statistically diminished depressive symptoms, which is also repeatedly shown in literature to be a beneficial effect of physical exercise in mood disorders [70].

The small sample of the passive group reassessed at 12 weeks presented no change in most of the parameters under investigation. They only showed an improvement in learning ability. This result may be explained by the short interval between the pre-post evaluation and familiarization with the evaluation process.

When participants were divided by genotype into two subgroups of APOE ε4 carriers and non-ε4 carriers, comparisons among the groups showed statistically significant differences only for the non-ε4 carriers' depressive symptoms assessed by GDS between the LLM and PTC groups, with a less depressive burden for the PTC group. The ε4 carriers did not differentiate for any parameter between the groups.

Within-group analysis, however, revealed statistically significant improvements in the LLM group in working memory and visuospatial ability as measured by the Trail B test for the ε4 carriers, while the non-ε4 carriers showed considerable improvements in multiple cognitive domains. No changes have been noticed for either of the subgroups regarding functionality and emotional burden (Multimedia Appendix 3 Figure iii).

The AC ε4 carrier group showed no significant improvement in cognitive, functional, or emotional status. The non-ε4 carrier subgroup revealed multiple domain improvement once again but also a mild worsening of their functionality assessed by FUCAS. This functional decline drives us to the assumption that physical activity helps the maintenance of complex functional tasks in comparison with cognitive training alone (Multimedia Appendix 3 Figure iv).

Finally, the PTC ε4 carrier group showed the only improvement in working memory while non-ε4 carriers had significant improvement in multiple domains (RAVLT1, CVLT total, CVLTD, digit span). The non-ε4 carriers in this group also had

slightly worse functionality assessed by the FUCAS. Both subgroups, however, showed a statistical improvement in depressive symptoms (Multimedia Appendix 3 Figure v).

APOE isoforms have been used as potential predictor markers for examining cognitive intervention effects in the current literature. Peter et al [71] suggest that the presence of APOE ε4 was not a significant predictor of any change in the cognitive variables, while another study underlines the potential beneficial effect of its absence [72]. The above findings are in line with ours, as we found stability in cognitive performance in the case of ε4 carriers and a significant improvement in the case of non-ε4 carriers.

The presence of APOE ε4 has also been related to biological and cognitive outcomes in physical exercise. Different brain metabolic responses to exercise related to the APOE isoforms have been reported [73], although their significance remains to be elucidated. Hence, Makino et al [74] suggest potential memory function benefits of physical exercise for ε4 carriers among older adults, also consistent with our findings, showing improvement in working memory in both groups that used physical training.

In our study, the APOE ε4 presence was related to a resistance of cognitive improvement, while the non-ε4 carriers showed multiple cognitive benefits. Although nonresponders, the ε4 carriers seemed to improve at least one of the test scores under investigation with combined cognitive and physical training or physical training via exergaming but not with cognitive training only. That may be explained by the beneficial effects of physical exercise activity on systematic and neurological biological parameters [67,69,73-76]. The question that stems from the above findings is whether this is an interventional failure or the success of the disease progression postponement. The answer to the above question will be presented in a future paper on this study.

Limitations

Our study has certain limitations. Randomization and blinding of test administrators and participants was not feasible due to practical issues. However, the lack of randomization is unlikely to bias effects as demographic characteristics and baseline performance are comparable. In our sample, 17.6% (43/244) were APOE ε4 carriers (at least one ε4 allele), which is lower than the expected Greek population frequency of 25.5% [77]. That is due to the small sample size (n=244) regarding genetic studies, although other cognitive intervention studies recruited even fewer participants (n<100 [72,73]).

The short time to reevaluation may influence the described effect on neuropsychological test scores. However, the short-term interval of 12 weeks for intervention is commonly used in the design of similar studies. The missing values issue is a common problem in clinical research. It was handled in the most effective and statistically approved way as described in the literature, considering all the parameters and running all the necessary tests to avoid statistical analyses bias. Regarding future review and meta-analysis, we should declare that 35 participants in the LLM group were part of the interventional group of a previous study [19]. Future research in the field should consider these difficulties and may overcome them by using larger samples and long-term follow-up.

Conclusions

Exergaming is an effective intervention method for patients with MCI. None of the applied computer-based interventions could be identified as the best. Nevertheless, it seems that combined cognitive and physical training and physical training via exergaming tend to be more effective for the high-risk MCI APOE ε4+ subgroup.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. However, parts of this work were supported by the LLM Care [58] self-funded initiative that emerged as the not-for-profit business exploitation of the Long-Lasting Memories (LLM Project) [78] originally funded by the ICT-CIP-PSP (Information and Communication Technologies–Competitiveness and Innovation Framework Programme–Policy Support Programme) Program of the European Commission.

Conflicts of Interest

PDB discloses potential (non-financial and beyond the context of the submitted work) conflicts of interest with PositScience: there is a co-marketing agreement between the company and the Aristotle University of Thessaloniki to exploit Brain HQ within the LLM Care self-funded initiative that emerged as the non-for-profit business exploitation of the Long-Lasting Memories (LLM Project) (www.longlastingmemories.eu) originally funded by the ICT-CIP-PSP Program of the European Commission. Brain HQ now forms part of LLM Care, a technology transfer/self-funded initiative that emerged as the non-for-profit business exploitation of LLM. FitForAll (FFA) has been developed in the Aristotle University of Thessaloniki during the Long-Lasting Memories (LLM Project) (www.longlastingmemories.eu) originally funded by the ICT-CIP-PSP Program of the European Commission. It now forms part of LLM Care, a technology transfer/self-funded initiative that emerged as the non-for-profit business exploitation of LLM.

Multimedia Appendix 1
Statistical analysis.

[DOCX File , 25 KB - jmir_v22i5e14617_app1.docx]

Multimedia Appendix 2
Supplementary tables.
[DOCX File , 33 KB - jmir_v22i5e14617_app2.docx]

Multimedia Appendix 3
Detailed results.
[DOCX File , 1334 KB - jmir_v22i5e14617_app3.docx]

References

1. Petersen RC. Mild cognitive impairment as a diagnostic entity. J Intern Med 2004 Sep;256(3):183-194 [FREE Full text] [doi: 10.1111/j.1365-2796.2004.01388.x] [Medline: 15324362]
2. Boyle PA, Buchman AS, Wilson RS, Kelly JF, Bennett DA. The APOE epsilon4 allele is associated with incident mild cognitive impairment among community-dwelling older persons. Neuroepidemiology 2010;34(1):43-49 [FREE Full text] [doi: 10.1159/000256662] [Medline: 19907191]
3. Belleville S, Fouquet C, Hudon C, Zomahoun HTV, Croteau J, Consortium for the Early Identification of Alzheimer's disease-Quebec. Neuropsychological measures that predict progression from mild cognitive impairment to Alzheimer's type dementia in older adults: a systematic review and meta-analysis. Neuropsychol Rev 2017 Dec;27(4):328-353 [FREE Full text] [doi: 10.1007/s11065-017-9361-5] [Medline: 29019061]
4. Counts SE, He B, Nadeem M, Wuu J, Scheff SW, Mufson EJ. Hippocampal drebrin loss in mild cognitive impairment. Neurodegener Dis 2012;10(1-4):216-219 [FREE Full text] [doi: 10.1159/000333122] [Medline: 22310934]
5. Sherman DS, Mauser J, Nuno M, Sherzai D. The efficacy of cognitive intervention in mild cognitive impairment (MCI): a meta-analysis of outcomes on neuropsychological measures. Neuropsychol Rev 2017 Dec;27(4):440-484 [FREE Full text] [doi: 10.1007/s11065-017-9363-3] [Medline: 29282641]
6. Bamidis PD, Vivas AB, Styliadis C, Frantzidis C, Klados M, Schlee W, et al. A review of physical and cognitive interventions in aging. Neurosci Biobehav Rev 2014 Jul;44:206-220. [doi: 10.1016/j.neubiorev.2014.03.019] [Medline: 24705268]
7. Coyle H, Traynor V, Solowij N. Computerized and virtual reality cognitive training for individuals at high risk of cognitive decline: systematic review of the literature. Am J Geriatr Psychiatry 2015 Apr;23(4):335-359. [doi: 10.1016/j.jagp.2014.04.009] [Medline: 24998488]
8. López-Higes R, Martín-Aragoneses MT, Rubio-Valdehita S, Delgado-Losada ML, Montejo P, Montenegro M, et al. Efficacy of cognitive training in older adults with and without subjective cognitive decline is associated with inhibition efficiency and working memory span, not with cognitive reserve. Front Aging Neurosci 2018;10:23 [FREE Full text] [doi: 10.3389/fnagi.2018.00023] [Medline: 29456502]
9. Miotto E, Batista A, Simon S, Hampstead B. Neurophysiologic and cognitive changes arising from cognitive training interventions in persons with mild cognitive impairment: a systematic review. Neural Plast 2018:1-4. [doi: 10.1155/2018/7301530]
10. Poptsi E, Lazarou I, Markou N, Vassiloglou M, Nikolaidou E, Diamantidou A, et al. A comparative single-blind randomized controlled trial with language training in people with mild cognitive impairment. Am J Alzheimers Dis Other Demen 2019 May;34(3):176-187. [doi: 10.1177/1533317518813554] [Medline: 30518237]
11. Vivas AB, Ypsilanti A, Ladas AI, Kounti F, Tsolaki M, Estévez AF. Enhancement of visuospatial working memory by the differential outcomes procedure in mild cognitive impairment and Alzheimer's disease. Front Aging Neurosci 2018;10:364 [FREE Full text] [doi: 10.3389/fnagi.2018.00364] [Medline: 30524264]
12. Tsolaki M, Kounti F, Agogiatou C, Poptsi E, Bakoglidou E, Zafeiropoulou M, et al. Effectiveness of nonpharmacological approaches in patients with mild cognitive impairment. Neurodegener Dis 2011;8(3):138-145. [doi: 10.1159/000320575] [Medline: 21135531]
13. Diederich K, Bastl A, Wersching H, Teuber A, Strecker J, Schmidt A, et al. Effects of different exercise strategies and intensities on memory performance and neurogenesis. Front Behav Neurosci 2017;11:47 [FREE Full text] [doi: 10.3389/fnbeh.2017.00047] [Medline: 28360847]
14. Brown BM, Sohrabi HR, Taddei K, Gardener SL, Rainey-Smith SR, Peiffer JJ, Dominantly Inherited Alzheimer Network. Habitual exercise levels are associated with cerebral amyloid load in presymptomatic autosomal dominant Alzheimer's disease. Alzheimers Dement 2017 Nov;13(11):1197-1206 [FREE Full text] [doi: 10.1016/j.jalz.2017.03.008] [Medline: 28501451]
15. García-Mesa Y, Colie S, Corpas R, Cristòfol R, Comellas F, Nebreda AR, et al. Oxidative stress is a central target for physical exercise neuroprotection against pathological brain aging. J Gerontol A Biol Sci Med Sci 2016 Jan;71(1):40-49. [doi: 10.1093/gerona/glv005] [Medline: 25720862]

16. Abd El-Kader SM, Al-Jiffri OH. Aerobic exercise improves quality of life, psychological well-being and systemic inflammation in subjects with Alzheimer's disease. Afr Health Sci 2016 Dec;16(4):1045-1055 [FREE Full text] [doi: 10.4314/ahs.v16i4.22] [Medline: 28479898]
17. Head D, Bugg JM, Goate AM, Fagan AM, Mintun MA, Benzinger T, et al. Exercise engagement as a moderator of the effects of APOE genotype on amyloid deposition. Arch Neurol 2012 May;69(5):636-643 [FREE Full text] [doi: 10.1001/archneurol.2011.845] [Medline: 22232206]
18. Erickson KI, Gildengers AG, Butters MA. Physical activity and brain plasticity in late adulthood. Dialogues Clin Neurosci 2013 Mar;15(1):99-108 [FREE Full text] [Medline: 23576893]
19. Bamidis PD, Fissler P, Papageorgiou SG, Zilidou V, Konstantinidis EI, Billis AS, et al. Gains in cognition through combined cognitive and physical training: the role of training dosage and severity of neurocognitive disorder. Front Aging Neurosci 2015;7:152 [FREE Full text] [doi: 10.3389/fnagi.2015.00152] [Medline: 26300772]
20. Konstantinidis EI, Bamparopoulos G, Bamidis PD. Moving real exergaming engines on the web: the webFitForAll case study in an active and healthy ageing living lab environment. IEEE J Biomed Health Inform 2017 Dec;21(3):859-866. [doi: 10.1109/JBHI.2016.2559787] [Medline: 28113566]
21. Kueider AM, Parisi JM, Gross AL, Rebok GW. Computerized cognitive training with older adults: a systematic review. PLoS One 2012;7(7):e40588 [FREE Full text] [doi: 10.1371/journal.pone.0040588] [Medline: 22792378]
22. Anderson-Hanley C, Maloney M, Barcelos N, Striegnitz K, Kramer A. Neuropsychological benefits of neuro-exergaming for older adults: a pilot study of an interactive physical and cognitive exercise system (iPACES™). J Aging Phys Act 2016 Jun 23;25:73-83. [doi: 10.1123/japa.2015-0261] [Medline: 27337738]
23. Konstantinidis EI, Billis AS, Mouzakidis CA, Zilidou VI, Antoniou PE, Bamidis PD. Design, implementation, and wide pilot deployment of FitForAll: an easy to use exergaming platform improving physical fitness and life quality of senior citizens. IEEE J Biomed Health Inform 2016 Jan;20(1):189-200. [doi: 10.1109/JBHI.2014.2378814] [Medline: 26731797]
24. Sun X, Nicholas J, Walker A, Wagner MT, Bachman D. APOE genotype in the diagnosis of Alzheimer's disease in patients with cognitive impairment. Am J Alzheimers Dis Other Demen 2012 Aug;27(5):315-320. [doi: 10.1177/1533317512452037] [Medline: 22815080]
25. Wolf AB, Caselli RJ, Reiman EM, Valla J. APOE and neuroenergetics: an emerging paradigm in Alzheimer's disease. Neurobiol Aging 2013 Apr;34(4):1007-1017 [FREE Full text] [doi: 10.1016/j.neurobiolaging.2012.10.011] [Medline: 23159550]
26. Kim J, Basak JM, Holtzman DM. The role of apolipoprotein E in Alzheimer's disease. Neuron 2009 Aug 13;63(3):287-303 [FREE Full text] [doi: 10.1016/j.neuron.2009.06.026] [Medline: 19679070]
27. O'Donoghue MC, Murphy SE, Zamboni G, Nobre AC, Mackay CE. APOE genotype and cognition in healthy individuals at risk of Alzheimer's disease: a review. Cortex 2018 Dec;104:103-123 [FREE Full text] [doi: 10.1016/j.cortex.2018.03.025] [Medline: 29800787]
28. Bonham LW, Geier EG, Fan CC, Leong JK, Besser L, Kukull WA, et al. Age-dependent effects of APOE ε4 in preclinical Alzheimer's disease. Ann Clin Transl Neurol 2016 Sep;3(9):668-677 [FREE Full text] [doi: 10.1002/acn3.333] [Medline: 27648456]
29. Harris AD, McGregor JC, Perencevich EN, Furuno JP, Zhu J, Peterson DE, et al. The use and interpretation of quasi-experimental studies in medical informatics. J Am Med Inform Assoc 2006;13(1):16-23 [FREE Full text] [doi: 10.1197/jamia.M1749] [Medline: 16221933]
30. González-Palau F, Franco M, Bamidis P, Losada R, Parra E, Papageorgiou SG, et al. The effects of a computer-based cognitive and physical training program in a healthy and mildly cognitive impaired aging sample. Aging Ment Health 2014 Sep;18(7):838-846. [doi: 10.1080/13607863.2014.899972] [Medline: 24697325]
31. Folstein MF, Folstein SE, McHugh PR. “Mini-mental state”: a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198. [Medline: 1202204]
32. Fountoulakis KN, Tsolaki M, Chantzi H, Kazis A. Mini Mental State Examination (MMSE): a validation study in Greece. Am J Alzheimers Dis Other Demen 2000 Nov 01;15(6):342-345. [doi: 10.1177/153331750001500604]
33. Kounti F, Tsolaki M, Eleftheriou M, Agogiatou C, Karagiozi K, Bakoglidou E. Administration of Montreal Cognitive Assesment (MoCA) test in Greek healthy elderly, patients with mild cognitive impairment and patients with dementia. 2007 Presented at: 2nd International Conference on Psychological Assessment; 2007; Thessaloniki.
34. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005 Apr;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]
35. Poptsi E, Kounti F, Karagiozi K, Eleuftheriou M, Agogiatou C, Bacoglidou E. The administration of trail making test to Greek healthy elderly, to patients with mild cognitive impairment, preclinical dementia and mild dementia. 2007 Presented at: 2nd International Conference on Psychological Assessment; 2007; Thessaloniki p. 157-158.
36. Armitage SG. An analysis of certain psychological tests used for the evaluation of brain injury. Psychological Monographs 1946;60(1):i-48. [doi: 10.1037/h0093567]
37. Delis D, Kramer J, Kaplan E, Ober B. California Verbal Learning Test: Adult Version. San Antonio: The Psychological Corporation; 1987.

38. Woods S, Delis D, Scott J, Kramer J, Holdnack J. The California Verbal Learning Test, second edition: test-retest reliability, practice effects, and reliable change indices for the standard and alternate forms. Arch Clin Neuropsychol 2006 Aug;21(5):413-420. [doi: 10.1016/j.acn.2006.06.002]
39. Vlahou CH, Kosmidis MH, Dardagani A, Tsotsi S, Giannakou M, Giazkoulidou A, et al. Development of the Greek Verbal Learning Test: reliability, construct validity, and normative standards. Arch Clin Neuropsychol 2012 Nov 23;28(1):52-64. [doi: 10.1093/arclin/acs099]
40. Schmidt M. Rey Auditory and Verbal Learning Test: A Handbook. Los Angeles: Western Psychological Services; 1996.
41. Messinis L, Tsakona I, Malefaki S, Papathanasopoulos P. Normative data and discriminant validity of Rey's Verbal Learning Test for the Greek adult population. Arch Clin Neuropsychol 2007 Aug;22(6):739-752. [doi: 10.1016/j.acn.2007.06.002]
42. Wechsler D. WMS-R: Wechsler Memory Scale-Revised manual. San Antonio: The Psychological Corporation; 1987.
43. Wechsler D. Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV). San Antonio: Pearson; 2014.
44. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist 1969;9(3):179-186. [Medline: 5349366]
45. Theotoka I, Kapaki E, Vagenas V, Ilias I, Paraskevas GP, Liappas I. Preliminary report of a validation study of Instrumental Activities of Daily Living in a Greek sample. Percept Mot Skills 2007 Jun;104(3 Pt 1):958-960. [doi: 10.2466/pms.104.3.958-960] [Medline: 17688151]
46. Hutton J, Dippel R, Loewenson R. Functional rating scale for the symptoms of dementia. In: Gallo J, Bogner H, Fulmer T, Paveza G, editors. Handbook of Geriatric Assessment. Rockville: Aspen Publishers; 1988.
47. Kounti F, Tsolaki M, Kiosseoglou G. Functional cognitive assessment scale (FUCAS): a new scale to assess executive cognitive function in daily life activities in patients with dementia and mild cognitive impairment. Hum Psychopharmacol 2006 Jul;21(5):305-311. [doi: 10.1002/hup.772] [Medline: 16856217]
48. Beck AT, Ward CH, Mendelson M, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry 1961 Jun;4:561-571. [doi: 10.1001/archpsyc.1961.01710120031004] [Medline: 13688369]
49. Tzemos I. The standardization of the Beck depression inventory in a Greek population sample [Dissertation]. Zografou: National Kapodistrian University of Athens; 1984.
50. Yesavage JA, Sheikh JI. Geriatric Depression Scale (GDS): recent evidence and development of a shorter version. Clin Gerontol 2008 Oct 25;5(1-2):165-173. [doi: 10.1300/J018v05n01_09]
51. Fountoulakis KN, Tsolaki M, Iacovides A, Yesavage J, O'Hara R, Kazis A, et al. The validation of the short form of the Geriatric Depression Scale (GDS) in Greece. Aging Clin Exp Res 2014 Jan 7;11(6):367-372. [doi: 10.1007/bf03339814]
52. Beck A, Steer R. Manual for the Beck Anxiety Inventory. San Antonio: Psychological Corporation; 1990.
53. Tselebis A, Gournas G, Tzitzanidou G, Panagiotou A, Ilias I. Anxiety and depression in Greek nursing and medical personnel. Psychol Rep 2006 Aug;99(1):93-96. [doi: 10.2466/pr0.99.1.93-96] [Medline: 17037454]
54. Salvi F, Miller M, Grilli A, Giorgi R, Towers A, Morichi V. A manual of guidelines to score the modified cumulative illness rating scale and its validation in acute hospitalized elderly patients. J Am Geriatr Soc 2008;56:31. [doi: 10.1111/j.1532-5415.2008.01935.x]
55. Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011 May;7(3):270-279 [FREE Full text] [doi: 10.1016/j.jalz.2011.03.008] [Medline: 21514249]
56. Billis A, Konstantinidis E, Mouzakidis C, Tsolaki M, Pappas C, Bamidis P. A game-like interface for training seniors' dynamic balance and coordination. In: Bamidis P, Pallikarakis N, editors. XII Mediterranean Conference on Medical and Biological Engineering and Computing 2010. Berlin: Springer; 2010.
57. Bamidis PD, Konstantinidis EI, Billis A, Frantzidis C, Tsolaki M, Hlauschek W, et al. A Web services-based exergaming platform for senior citizens: the Long Lasting Memories project approach to e-health care. Conf Proc IEEE Eng Med Biol Soc 2011;2011:2505-2509. [doi: 10.1109/IEMBS.2011.6090694] [Medline: 22254850]
58. Video GRade. URL: http://www.llmcare.gr [accessed 2019-07-31]
59. Koutroumani M, Daniilidou M, Giannakouros T, Proitsi P, Liapi D, Germanou A, et al. The deletion variant of α2b-adrenergic receptor is associated with decreased risk in Alzheimer's disease and mild cognitive impairment. J Neurol Sci 2013 May 15;328(1-2):19-23. [doi: 10.1016/j.jns.2013.02.003] [Medline: 23499426]
60. Tabachnick B, Fidell L. Using Multivariate Statistics. San Antonio: Pearson Education; 2013.
61. Dong Y, Peng CJ. Principled missing data methods for researchers. Springerplus 2013 Dec;2(1):222 [FREE Full text] [doi: 10.1186/2193-1801-2-222] [Medline: 23853744]
62. Little R, Schenker N. Missing Data. In: Arminger G, Clogg C, Sobel M, editors. Handbook of Statistical Modeling for the Social and Behavioral Sciences. Boston: Springer; 1995:39-75.
63. IBM Corp. 2011. IBM SPSS Missing Values 20 URL: ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/20.0/en/client/Manuals/IBM_SPSS_Missing_Values.pdf [accessed 2020-01-19]
64. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009;338:b2393 [FREE Full text] [Medline: 19564179]

65. van Ginkel JR, Kroonenberg PM. Analysis of variance of multiply imputed data. Multivariate Behav Res 2014 Jan 01;49(1):78-91 [FREE Full text] [doi: 10.1080/00273171.2013.855890] [Medline: 24860197]
66. Klimova B, Maresova P. Computer-based training programs for older people with mild cognitive impairment and/or dementia. Front Hum Neurosci 2017;11:262 [FREE Full text] [doi: 10.3389/fnhum.2017.00262] [Medline: 28559806]
67. Train the Brain Consortium. Randomized trial on the effects of a combined physical/cognitive training in aged MCI subjects: the Train the Brain study. Sci Rep 2017 Jan 03;7:39471 [FREE Full text] [doi: 10.1038/srep39471] [Medline: 28045051]
68. Kato T, Kanemaru A, Sugawara Y, Kawaji Y, Hiraoka T, Honda T, et al. A combination intervention (cognitive training and physical exercise) could improve or maintain cognitive functioning in MCI subjects. J Neurol Sci 2017 Oct;381:668. [doi: 10.1016/J.JNS.2017.08.1881]
69. Anderson-Hanley C, Barcelos NM, Zimmerman EA, Gillen RW, Dunnam M, Cohen BD, et al. The aerobic and cognitive exercise study (ACES) for community-dwelling older adults with or at-risk for mild cognitive impairment (MCI): neuropsychological, neurobiological and neuroimaging outcomes of a randomized clinical trial. Front Aging Neurosci 2018;10:76 [FREE Full text] [doi: 10.3389/fnagi.2018.00076] [Medline: 29780318]
70. Hearing CM, Chang WC, Szuhany KL, Deckersbach T, Nierenberg AA, Sylvia LG. Physical exercise for treatment of mood disorders: a critical review. Curr Behav Neurosci Rep 2016 Dec;3(4):350-359 [FREE Full text] [doi: 10.1007/s40473-016-0089-y] [Medline: 28503402]
71. Peter J, Schumacher LV, Landerer V, Abdulkadir A, Kaller CP, Lahr J, et al. Biological factors contributing to the response to cognitive training in mild cognitive impairment. J Alzheimers Dis 2018;61(1):333-345 [FREE Full text] [doi: 10.3233/JAD-170580] [Medline: 29154279]
72. Franzmeier N, Unterauer E. Effects of age, APOE ε4, cognitive reserve and hippocampal volume on cognitive intervention outcome in amnestic mild cognitive impairment. J Alzheimers Dis Parkinsonism 2016;6(4):1. [doi: 10.4172/2161-0460.1000246]
73. Porto FH, Coutinho AM, Duran FL, Ono CR, Pinto AL, Gualano B, et al. APOE genotype modifies the effects of aerobic training on brain glucose metabolism in subjects with mild cognitive impairment. Alzheimers Dement 2016 Jul;12(7):P588-P589. [doi: 10.1016/j.jalz.2016.06.1156]
74. Makino T, Umegaki H, Hayashi T, Cheng X, Shimada H, Dodge H, et al. [P2-575]: APOE EPSILON 4 STATUS MODIFIES THE EFFECTS OF AEROBIC, RESISTANCE, OR COMBINED TRAINING ON COGNITIVE FUNCTION IN OLDER ADULTS. Alzheimer's & Dementia 2017 Jul 01;13(7S_Part_17):P867-P867. [doi: 10.1016/j.jalz.2017.06.1234]
75. Lerner Y, Yogev-Seligmann G, Eisenstein T, Ash E, Hendler T, Giladi N. Brain plasticity following physical training in individuals with mild cognitive impairment: neuroimaging study. Alzheimers Dement 2017 Jul;13(7):P867-P868. [doi: 10.1016/j.jalz.2017.06.1235]
76. Fissler P, Küster O, Schlee W. Novelty interventions to enhance broad cognitive abilities and prevent dementia. Prog Brain Res 2013;207:403-434. [doi: 10.1016/b978-0-444-63327-9.00017-5] [Medline: 24309264]
77. Tsolaki AC, Gatzima O, Daniilidou M, Lazarou E, Bamidis PD, Verykouki E, et al. Prevalence of apolipoprotein E polymorphisms in Alzheimer's disease, mild cognitive impairment, and healthy elderly: a Northern Greece study. Neurodegener Dis 2018;18(4):216-224. [doi: 10.1159/000491764] [Medline: 30205398]
78. Long-Lasting Memories: mind and body fitness for life. URL: http://www.longlastingmemories.eu [accessed 2020-01-19]

Abbreviations

AC: Active Control
APOE: apolipoprotein E
APOE ε4: ε4 allele of the apolipoprotein gene
BAI: Beck Anxiety Inventory
BDI: Beck Depression Inventory
CIRS: Cumulative Illness Rating Scale–Geriatric
CVLT: California Verbal Learning Test
FFA: FitForAll
FRSSD: Functional Rating Scale of Symptoms of Dementia
FUCAS: Functional and Cognitive Assessment Test
GDS: Geriatric Depression Scale
IADL: Instrumental Activities of Daily Living scale
ICT: information and communication technologies
LLM: Long-Lasting Memories
MCI: mild cognitive impairment
MMSE: Mini-Mental State Examination
MoCA: Montreal Cognitive Assessment
PTC: Physical Training Control
RAVLT: Rey Auditory Verbal Learning Test

Trail B: Trail Making Test, Part B

Edited by G Eysenbach; submitted 17.07.19; peer-reviewed by M Tsatali, F Lanfranchi; comments to author 27.08.19; revised version received 19.10.19; accepted 15.12.19; published 07.05.20.

Please cite as:
Tsolaki AC, Tsolaki M, Pandria N, Lazarou E, Gkatzima O, Zilidou V, Karagianni M, Iakovidou-Kritsi Z, Kimiskidis VK, Bamidis PD
Web-Based Intervention Effects on Mild Cognitive Impairment Based on Apolipoprotein E Genotype: Quasi-Experimental Study
J Med Internet Res 2020;22(5):e14617
URL: https://www.jmir.org/2020/5/e14617
doi:10.2196/14617
PMID:32379048

Original Paper

Use of Brief Messages Based on Behavior Change Techniques to Encourage Medication Adherence in People With Type 2 Diabetes: Developmental Studies

Yvonne Kiera Bartlett[1], PhD; Andrew Farmer[2], DM, FRCGP; Rustam Rea[3,4], DM, FRCP; David P French[1], PhD

[1]Manchester Centre for Health Psychology, University of Manchester, Manchester, United Kingdom

[2]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

[3]Oxford Centre for Diabetes, Endocrinology and Metabolism, Oxford University Hospitals Foundation Trust, Oxford, United Kingdom

[4]National Institute for Health Research Oxford Biomedical Research Centre, Oxford University Hospitals National Health Service Foundation Trust, Oxford, United Kingdom

Corresponding Author:
Yvonne Kiera Bartlett, PhD
Manchester Centre for Health Psychology
University of Manchester
Coupland 1 Building
Oxford Road
Manchester
United Kingdom
Phone: 44 161 306 5436
Email: kiera.bartlett@manchester.ac.uk

Abstract

Background: Brief messages are a promising way to improve adherence to medication for people with type 2 diabetes. However, it is often unclear how messages have been developed and their precise content, making it difficult to ascertain why certain messages are successful and some are not.

Objective: The goal of the research was to develop messages that have proven fidelity to specified evidence-derived behavior change techniques (BCTs) and are acceptable to people with type 2 diabetes.

Methods: Four studies were conducted: (1) a workshop (n=21) where behavioral change researchers and health care professionals developed messages based on specific BCTs or beliefs or concerns related to taking medication, (2) a focus group study with people with type 2 diabetes (n=23) to assess acceptability of the approach, (3) a survey to ascertain the acceptability of a subset of messages to people with type 2 diabetes (n=61) and, (4) a survey with behavior change researchers to assess the fidelity of a subset of messages to their intended BCT (n=18).

Results: In study 1, 371 messages based on 38 BCTs and beliefs/concerns were developed. Workshop participants rated BCTs to be relevant to medication adherence (mean 7.12/10 [SD 1.55]) and messages to have good fidelity (mean 7.42/10 [SD 1.19]). In study 2, the approach of providing medication adherence support through text messages was found to be acceptable. In study 3, mean acceptability of all BCTs was found to be above the midpoint (mean 3.49/5 [SD 0.26]). In study 4, mean fidelity for all BCTs was found to be above the midpoint (mean 7.61/10 [SD 1.38]).

Conclusions: A library of brief messages acceptable to people with type 2 diabetes and representative of specific evidence-derived BCTs was developed. This approach allowed brief messages to be developed with known content that can be used to test theory.

(J Med Internet Res 2020;22(5):e15989) doi:10.2196/15989

KEYWORDS

behavior change; behavior change techniques; type 2 diabetes, brief messages; mHealth; medication adherence

Introduction

Type 2 diabetes leads to high levels of glucose in the blood; if left uncontrolled, the condition can lead to a wide range of micro- and macrovascular complications including problems with the heart, feet, and eyesight [1]. There are effective tablet medications available that can help control blood glucose and reduce the risk of complications in people with type 2 diabetes. However, adherence to and persistence with tablet medication is poor for at least half the population of people with type 2 diabetes prescribed these medications [2]. If less than 80% of the prescribed medication is taken, only half the expected reduction in blood glucose control is seen, increasing the risk of complications for this population [3].

A recent Cochrane review concluded that medication adherence interventions have led to only modest increases in adherence, and studies with the lowest risk of bias had the lowest efficacy [4]. The authors proposed that novel approaches to this problem are needed [4]. One approach that has shown promise to improve medication adherence in people with type 2 diabetes is the delivery of brief messages. In a review of 15 interventions, there was some evidence of effectiveness, but there were weaknesses in the study designs [5]. Further, the review authors commented that the interventions were not well described, and the majority did not report being based on any explicit theoretical framework [5]. It has been noted that for brief message interventions "the content is the central driver of the behavior change" [6]; it is therefore essential that the content is well described so that whether a change in behavior does or does not occur it can be explained, and the intervention can be improved as needed. However, in many published reports, in addition to the content of messages being unclear, the development process can also be unclear with little reported about exactly how information from formative work is incorporated into the final intervention. This lack of transparency has led to the development process being described as a black box [7].

There are multiple ways to develop interventions to change behavior, and different approaches may be suitable for different contexts [8]. These approaches vary in the emphasis they place on the extent to which interventions should be based on theory and evidence of effectiveness and whether they should be developed and/or extensively tested with the intended target population. More specifically in the electronic health area, there are a growing number of approaches that focus on developing technology-based interventions (eg, the person-centered approach [9], behavioral intervention technology model [10], intervention mapping combined with the behavioral intervention technology model [11], and the mobile health development and evaluation framework [12]). However, these models provide a broad overview for development of many different types of technology-based interventions, which are often complex with multiple components. For this reason, the actual process of writing content such as brief messages is not often explained in detail. In cases where the actual process of writing the messages is reported or is the focus of the suggested procedure [13], it seems that it is conducted by a small number of people in the intervention team. Having a small team of people involved in the intervention drafting the messages may lead to a similarity in the approaches used compared with what has gone before rather than exploring novel approaches to the problem of medication adherence.

This research developed and refined message content using specific behavior change techniques (BCTs) to develop messages or as a means to code messages for content. BCTs are described as the active ingredients of an intervention [14]. They are the irreducible components of the intervention such as goal-setting or self-monitoring of behavior. Thus far, 93 such BCTs have been identified and defined in a taxonomy [15]. In cases where the BCT content of brief messages related to diabetes self-management have been defined in published reports, thus far relatively few of the 93 BCTs identified in the v1 taxonomy [15] have been used. In two reviews in the area, one identified 8 BCTs across 7 studies [16] and the other 16 BCTs across 6 studies [17].

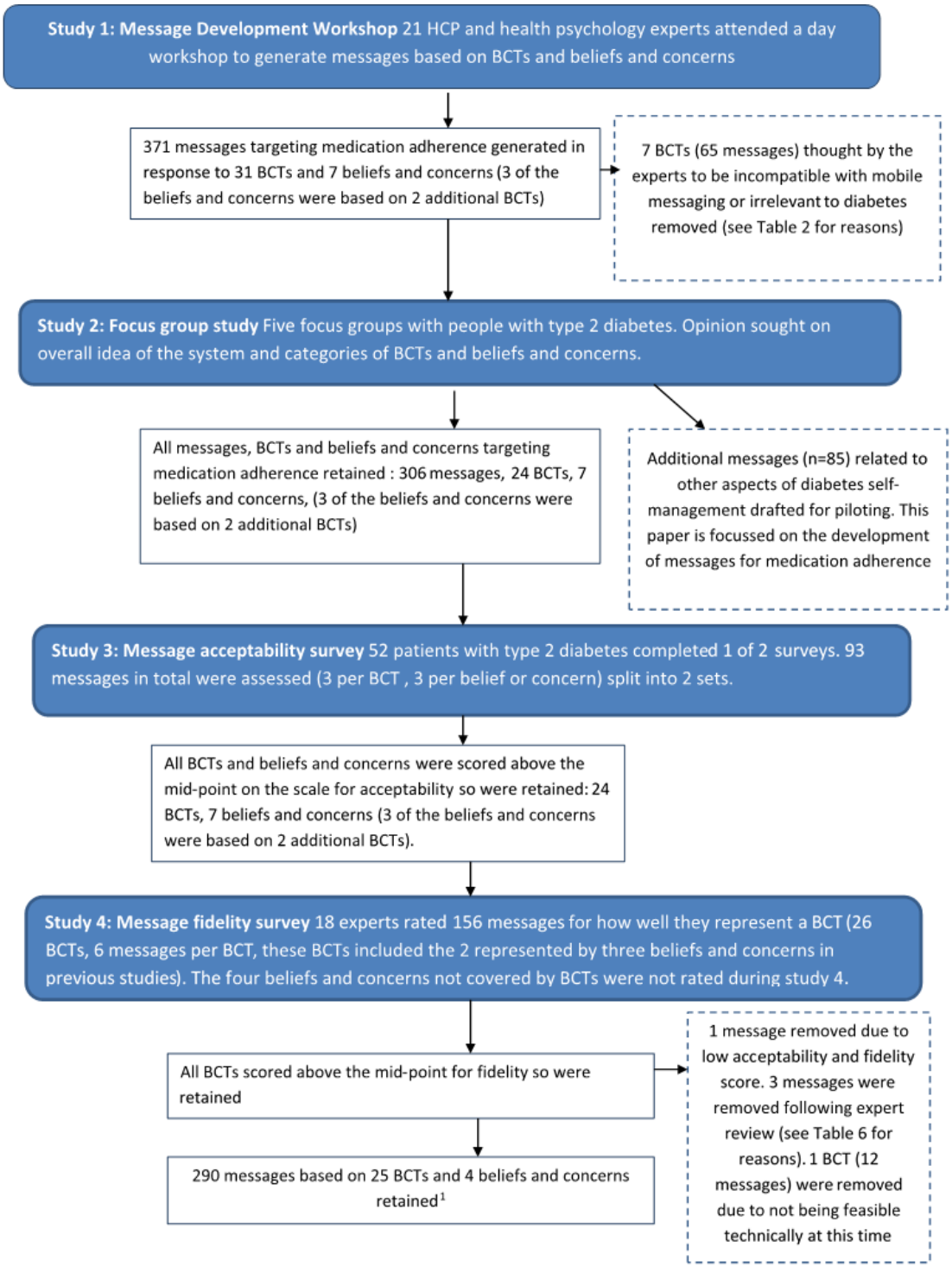
The studies described in this paper form part of a larger body of work to develop a brief message service that could be used by people with type 2 diabetes within the UK National Health Service (NHS). Given the lack of effectiveness of existing interventions to promote medication adherence [4], it was important to identify as many potentially effective BCTs that could be delivered by brief message as possible. Prior to the work reported here, a rapid systematic review of systematic reviews was conducted [18]. This review highlighted a wide range of BCTs that existing quantitative systematic reviews had previously identified as being associated with changes in medication adherence across a range of chronic physical health conditions. In addition, qualitative reviews were examined to identify specific beliefs and concerns associated with taking medication for people with type 2 diabetes to ensure that novel insights inductively obtained from people with type 2 diabetes regarding their medication adherence were included [18]. Of the 46 BCTs identified, 13 were either not considered suitable for delivery via brief messages or, in delivering through brief messages, the content could be covered by another BCT (for excluded BCTs and reasons see Multimedia Appendix 1, Table A), 2 of the BCTs were considered more appropriately represented by 3 of the beliefs and concerns (see Table A for details).

This paper presents a transparent process of brief message development that aimed to ensure the message content had high fidelity to the intended BCTs. Further, the inclusion of relevant stakeholders at all stages of the research was deemed essential to ensure that the messages were acceptable to the target population. People with type 2 diabetes were included both as participants in the research studies but also in Patient and Public Involvement (PPI) panels. Members of the PPI panel participated in key meetings and research decisions, as well as reviewing protocols and participant-facing materials prior to ethical approval.

This paper reports the process of message development through 4 studies and decisions taken following each study. The purpose of this research was to develop a library of messages that are acceptable and represent explicit BCTs, identify any BCTs that were not suitable for delivery through brief messages, and identify any messages or BCTs that should be removed prior

to further development. The studies reported on are (1) a workshop to develop a library of messages based on specified BCTs and beliefs and concerns, (2) a focus group study to look at acceptability of the concept, (3) an online survey to assess the acceptability of the BCTs and messages developed during study 1 to people with type 2 diabetes, and (4) an online survey to assess the acceptable messages from study 3 for fidelity to the intended BCTs (see Figure 1 for an overview of message development through the four studies).

Figure 1. Development of messages targeting medication adherence through the four studies. BCT: behavior change technique; [1]Three of the beliefs and concerns are represented by 2 additional BCTs here.



Methods

Ethical Approval

Ethical approval for the studies 1 and 4 was granted by the University of Oxford Central University Research Ethics Committee Medical Sciences Interdivisional Research Ethics Committee (ref: R50752/RE001). Ethical approval for studies 2 and 3 was granted by the NHS North West–Greater Manchester West Research Ethics Committee (ref: 17/NW/0224).

Study 1: Message Development Workshop

Aim

The aim of the workshop was to develop a library of messages based on specified BCTs and beliefs and concerns.

Recruitment

Participants were expert researchers in the fields of behavior change or medication adherence or health care professionals (HCPs) involved in diabetes care. Potential participants were identified from the research team's knowledge of those active in these fields in the United Kingdom or Europe. A total of 31 invitations were sent by email, with a single reminder email.

Procedure

Participants were invited to attend a 1-day workshop in Manchester, England. An information sheet was sent to participants ahead of time and informed consent was taken on the day of the workshop. The day was split into 4 sessions. For the first 3 sessions, behavior change researchers were split into 4 groups of 4 or 5 people and for each session were provided with the descriptions of 2 or 3 BCTs per group. HCPs were asked to generate messages to address specific beliefs and concerns. The 31 BCTs and 7 beliefs and concerns presented in study 1 were those identified by rapid review [18] and considered to be plausible for delivery via brief messages. Three of the included beliefs and concerns presented to HCPs represented an additional 2 BCTs. "Difficulties remembering and understanding the medication regimen" represented the BCT 4.1: Instruction on how to perform a behavior and "perceived risks of taking medication" and "beliefs about medication necessity" represented BCT 5.1: Information about health consequences (Table 1). Each group had a moderator who did not actively participate in writing messages but asked participants to develop brief messages based on each BCT that would target tablet medication adherence for people with diabetes. The moderator then entered any messages that were agreed by the group onto a spreadsheet. The messages generated by behavior change researchers were reviewed by a different group of behavior change researchers and rated according to how relevant the BCT was thought to be to improving diabetes tablet medication adherence (on a scale of 1 to 10), how far the aim of getting 8 to 10 messages that reflect the BCT well had been achieved (scale of 1 to 10), and the fidelity of each message generated to the intended BCT (scale of 1 to 10). The group of HCPs remained as a single group throughout and rated their own messages for how well they thought the messages addressed the beliefs and concerns raised in the rapid review, how relevant the belief/concern was thought to be, and how well the aim of getting 8 to 10 messages that reflect the belief/concern well was reached (scale of 1 to 10).

In the fourth session, the ratings for how well each BCT or belief and concern met the aim of generating 8 to 10 good quality messages were ranked, and those that scored lowest were discussed with all participants as a single group. Suitability of the BCTs for delivery through text message was discussed. Participants were asked to complete a survey describing their level of expertise and a brief workshop evaluation form. Participants were given an honorarium for their time, and accommodation, travel, and other subsistence costs were reimbursed.

Following the workshop, only those BCTs that were deemed compatible with delivery by text message and suitable for medication adherence for people with type 2 diabetes with an appropriate score for fidelity to the BCT were retained and used in study 2.

Study 2: Focus Group Study

Aim

The aim of the research was to ascertain the acceptability of the concept and types of messages to people with type 2 diabetes.

Recruitment

Five general practices sent letters introducing the study to eligible patients. Participants were eligible if they were aged over 18 years, taking tablet medication for type 2 diabetes, and had access to a mobile phone. Those who had been hospitalized in the last 3 months for hypo- or hyperglycemia, were pregnant or within 3 months postpartum, or had been diagnosed with a terminal illness were excluded.

Procedure

Participants were invited to take part in a focus group at the university or in a community location. Researchers facilitated the focus groups and presented the 31 BCTs that remained following study 1. As it would not be possible to discuss all 31 BCTs, these were grouped as potential strategies for behavior change grouped according to v1 taxonomy categories [15] with sample messages. As an example, BCT 1.2: Problem solving and BCT 1.4: Action planning are both grouped under goals and planning in the BCT v1 taxonomy [15]. Participants were given a description of the group "Messages that support and encourage you to plan to take your medication as intended, along with messages that encourage you to solve problems to help you achieve those goals" and some sample messages from those generated during study 1 were given (eg, "Plan when, where, and how you are going to take your medication" [BCT 1.4]).

The task of the groups was to discuss wording and acceptability of the idea of sending brief messages to encourage and support medication adherence overall rather than determine acceptability of individual messages or BCTs (the examples were given to provide a concrete grounding for discussion). Focus groups were recorded, transcribed, and analyzed thematically.

Following the focus group study, suggestions made by participants were reviewed and acted on as appropriate.

Study 3: Message Acceptability Survey

Aim

The goal of the research was to assess the acceptability of the BCTs and messages developed during studies 1 and 2 to people with type 2 diabetes.

Recruitment

Participants were aged over 18 years with type 2 diabetes and were taking tablet medication to control their diabetes (with or without concomitant insulin). Potential participants were invited from a nationwide database of individuals facilitated by the Greater Manchester Clinical Research Network. They had been diagnosed with diabetes, were interested in taking part in research, and had indicated on signing up to the database that they would be interested in completing online questionnaires (n=861). Advertisements were also posted on the Diabetes UK online support forum, and information about the study was distributed at a diabetes support group.

Procedure

The database facilitators sent initial email invitations containing a link to the participant information sheet, an online consent form, and 1 of 4 versions of the survey (see survey development and content section). Participants were screened for eligibility and required to agree with all consent statements before they could proceed to the survey. The advertisement also contained contact details if potential participants preferred to receive a paper version of the survey. In this case, the information sheet, consent forms, and 1 of the 4 surveys were printed and mailed to the individual with two postage paid return envelopes to keep consent forms and research data separate. No incentive was offered for completion.

Survey Development and Content

Two sets of messages from those developed in studies 1 and 2 were assessed for acceptability; 47 in survey 1, and 46 in survey 2. Messages were not grouped or ordered by BCT or belief/concern; both surveys were presented with messages in one order, and the reversed order, resulting in 4 versions of the survey. Participants were allocated sequentially to one of the versions of the survey. In total, 93 messages across the 31 BCTs and beliefs/concerns that remained after studies 1 and 2 were assessed (3 messages per BCT and belief/concern). Participants were asked to provide ratings for individual messages, and acceptability of each BCT or belief/concern was then considered across the 3 messages. The individual messages were selected by ordering per BCT or belief/concern according to the mean fidelity to the BCT or belief/concern score given to each message during study 1. For each BCT or belief/concern, one message that scored high, one that scored low, and one that was scored in the middle were chosen. This allowed examination of the acceptability across the range of fidelity to each BCT or belief/concern. Participants were asked initial screening questions, demographic questions, and whether they used a phone or tablet computer. Participants were then presented with either the 46 or 47 messages and after each message asked to provide 3 ratings of acceptability on 5-point Likert scales: How easy is the message to understand (very difficult to understand, difficult to understand, neither easy nor difficult to understand, easy to understand, very easy to understand)? How much do you like the message (do not like at all, do not like, neither like nor do not like, like, like a lot)? How useful would this message be for you (not useful at all to me, not useful to me, neither useful nor not useful to me, useful to me, very useful to me)? The final question was an open-ended comment box for any general comments about the messages. A panel of PPI representatives reviewed the survey, and changes were made in line with their suggestions prior to the survey being sent out. PPI representatives were reimbursed in accordance with the Involve guidelines from the National Institute of Health Research [19].

Analysis

For each facet of acceptability, individual message scores were examined to identify any messages that should be removed from the library. Mean scores for ease of understanding, liking, and usefulness were calculated across the messages for each BCT or belief/concern. An overall acceptability score was calculated as an average of the understanding, liking, and usefulness scores for each BCT or belief/concern.

Following the survey, only messages and BCTs, beliefs, and concerns that scored adequately in terms of acceptability were retained for study 4.

Study 4: Message Fidelity Survey

Aim

The goal of the research was to assess the extent to which the messages with any amendments following studies 2 and 3 had fidelity to the intended BCTs.

Recruitment

Participants were expert researchers in behavior change. Potential participants were identified from those who were invited but unable to attend study 1 and the research team's knowledge of those researching BCTs. Initial email invitations (n=32) were sent, and a single reminder email was sent to those who did not respond.

Procedure

If an individual responded positively to the invitation, they were sent a further email with a unique participant number, a link to the full participant information sheet, consent form, and 1 of 2 versions of the survey selected at random (one with the order of BCTs reversed). A random allocation sequence was generated in blocks of 10 using QuickCalcs (GraphPad Software), and the unique participant number was assigned the next allocation in the sequence. Participants had to agree to all consent statements before continuing to the survey. Participants were given an honorarium for completing the survey.

Survey Development and Content

Participants completed some brief questions about their expertise. For each BCT (n=26), participants were presented with the title and description of the BCT from the BCT v1 taxonomy [15], and then 6 sample messages were presented

with "How well does this message reflect the BCT *problem solving* as defined above?" under each message (italics adjusted for each BCT). Answers were given on a 10-point scale anchored with 1 (not very well) to 10 (very well). As participants were behavior change researchers, ratings were not sought for the beliefs and concerns but rather on whether or not the message reflected an underlying BCT. Three of the beliefs and concerns represented an additional 2 BCTs, resulting in 26 BCTs (eg, messages developed for the concern "perceived risks of taking medication" were presented under BCT 5.1: Information about health consequences). See Table 1 for details of beliefs and concerns and associated BCTs. For each of the 26 BCTs, 6 messages were rated, totaling 156 messages. The 6 messages chosen were those rated as having the highest fidelity to the BCT during study 1. The final question was an open-ended comment box for any general comments about the messages.

Results

Study 1: Message Development Workshop

Participants

A total of 21 participants attended the workshop. Participants were employed as either researchers (13/21, 62%), both researchers and HCPs (7/21, 33%), or as HCPs (1/21, 5%). Participants had been paid to conduct research and/or health care professional work for between 6 and 40 years (mean 15.95 [SD 9.94]). All participants (n=21) described behavior change interventions as either central or somewhat central to their work, 62% (13/21) described medication adherence as central or somewhat central, and 43% (9/20; 1 missing) of participants described diabetes as central or somewhat central to their work. Participants had published between 2 and 50 papers (mean 14.82 [SD 13.47]) related to medication adherence, diabetes or behavior change interventions (19/21, 1 missing and 1 clinician marked N/A).

Behavior Change Techniques

A total of 371 messages were generated during the workshop; 356 were scored and 15 were generated as additional suggested messages by the participants when they had finished scoring. There were between 6 and 15 messages per BCT or belief/concern (mean 9.76 [SD 2.07]). All subsequent BCT codes reference the v1 BCT taxonomy [15]. On a scale of 1 to 10, perceived relevance of the BCT or belief/concern to tablet medication adherence for people with type 2 diabetes ranged from 4.00 (BCT 13.1: Identification of self as a role model) to 9.75 (BCT 8.3: Habit formation) with a median of 8.00. When asked the extent to which the aim of writing 8 to 10 messages that reflect the BCT or belief/concern well had been achieved (from 1=not at all achieved to 10=completely achieved), the mean scores across reviewers ranged from 3.25 (BCT 8.4: Habit reversal) to 9.25 (BCT 5.3: Information about social and environmental consequences) with a median of 7.00. Across the messages, each BCT or belief/concern received mean fidelity scores ranging from 3.50 (BCT 8.4: Habit reversal) to 8.58 (BCT 3.1 Social support [unspecified]) with a median of 8.00. See Table 1 for details.

Decisions for Message Development

Following discussions in the fourth session of the workshop, 7 BCTs were identified as either incompatible with delivery via brief message during the workshop or unsuitable for medication adherence for people with type 2 diabetes: BCT 8.4: Habit reversal; BCT 11.3: Conserving mental resources; BCT 12.2: Restructuring the social environment; BCT 13.1: Identification of self as a role model; BCT 13.2 Framing/reframing; BCT 13.3: Incompatible beliefs; and BCT 15.3: Focus on past success. For detailed reasons, see Table 2. Following study 1, the library contained 306 messages based on 24 BCTs and 7 beliefs and concerns (3 of these beliefs and concerns were based on an additional 2 BCTs).

Study 2: Focus Group Study

Participants

A total of 23 participants with a mean age of 68 (SD 7.25) years and a mean of 11 (SD 5.94) years since diagnosis with type 2 diabetes took part in 1 of 5 focus groups; 17% (3/23) of the participants were female and 83% (15/23) were white British.

Behavior Change Techniques

Participants found the concept of sending messages related to medication adherence for diabetes acceptable, provided the messages were novel (eg, not using strategies the individual was already using). Participants were keen to introduce messages related to aspects of diabetes self-management beyond medication adherence and had suggestions for changes to the wording of some messages. The wider findings from these focus groups relating to the acceptability to people with type 2 diabetes of receiving short message service (SMS) messages in principle and what features the system should have are reported elsewhere [20].

Decisions for Message Development

Following the focus groups, suggestions were identified from the transcripts, and each was considered by the research team according to criteria adapted from previous research [21]. The criteria used to aid discussion were (1) how relevant the suggestion is to the target behavior, (2) how available the suggested content already is (if already widely available perhaps not needed in present intervention), (3) the ease of implementation from a technical perspective, and (4) support from theoretical findings and/or evidence. One researcher applied the criteria (with + for support, – for no support, and ? if unsure for each criteria on each suggestion); this was then brought to a multidisciplinary team meeting, agreements and disagreements about the rating were discussed, and a final decision was taken to accept or reject each suggestion. Following these discussions, messages related to other aspects of diabetes self-management were added in preparation for piloting (eg, links to verified sources of diet and exercise advice), and messages were reworded to use the word tablets rather than meds, medication, or pills in accordance with participants' preference. The number of messages targeting medication adherence remained at 306, based on 24 BCTs and 7 beliefs and concerns (3 of these beliefs and concerns were based on an additional 2 BCTs).

Table 1. Number of messages, mean perceived relevance of the behavior change techniques (BCTs) to medication adherence, mean score for how well the aim of 8 to 10 messages that reflected the BCT well was met, and mean fidelity of messages to the intended BCT or belief/concern from study 1. BCTs are identified through the codes from BCT v1 taxonomy [15] and beliefs and concerns are labeled A through G.

BCT[a] or belief/concern	Number of messages	Relevance to improving medication adherence for people with type 2 diabetes, mean (SD)	Aim to have 8 to 10 messages that reflect the BCT or belief/concern well, mean (SD)	Fidelity of messages to the intended BCT/belief or concern[b], mean (SD)
1.2. Problem solving	10	8.00 (1.41)	7.00 (0)	7.18 (1.36)
1.4. Action planning	8	8.50 (0.71)	7.25 (0.96)	7.94 (1.92)
2.3. Self-monitoring of behavior	11	9.00 (1.55)	6.75 (0.96)	6.86 (1.94)
3.1. Social support (unspecified)	9	9.25 (0.96)	9.00 (0)	8.58 (1.40)
3.2. Social support (practical)	8	9.00 (0.82)	8.50 (1.29)	7.83 (1.42)
3.3. Social support (emotional)	9	7.67 (1.15)	7.50 (1.00)	7.72 (1.75)
4.2. Information about antecedents	6	8.50 (1.29)	8.75 (0.96)	8.17 (1.34)
5.3. Information about social and environmental consequences	13	7.50 (1.91)	9.25 (0.50)	8.50 (1.13)
5.5. Anticipated regret	11	7.50 (2.38)	7.50 (0.58)	7.34 (1.36)
5.6. Information about emotional consequences	7	7.67 (1.15)	7.75 (0.50)	7.36 (1.91)
6.2. Social comparison	11	6.00 (2.45)	7.75 (0.96)	6.85 (2.63)
6.3. Information about others' approval	8	5.50 (3.54)	7.75 (1.71)	8.09 (1.84)
7.1. Prompts/cues	11	8.25 (0.96)	8.25 (0.50)	8.02 (1.34)
8.3. Habit formation	11	9.75 (0.50)	8.50 (1.00)	7.81 (1.48)
8.4. Habit reversal	3	4.50 (1.73)	3.25 (0.50)	3.50 (0.90)
9.1. Credible source	8	7.25 (2.22)	7.75 (0.50)	8.06 (1.64)
9.2. Pros and cons	11	6.67 (0.58)	8.25 (0.96)	8.31 (1.15)
9.3. Comparative imaginings of future outcomes	8	8.00 (1.00)	6.75 (0.50)	6.61 (1.99)
10.5. Social reward	12	7.50 (0.58)	8.50 (0.58)	7.71 (1.35)
11.2. Reduce negative emotions	8	6.50 (1.29)	6.75 (0.50)	6.97 (1.11)
11.3. Conserving mental resources	12	6.00 (2.16)	5.75 (1.26)	6.33 (1.02)
12.1. Restructuring the physical environment	13	9.00 (0.82)	8.50 (1.29)	8.25 (1.57)
12.2. Restructuring the social environment	9	6.00 (2.16)	4.75 (1.89)	7.00 (1.78)
13.1. Identification of self as role model	9	4.00 (2.45)	8.00 (0.82)	7.81 (2.14)
13.2. Framing/reframing	8	5.00 (1.00)	6.75 (0.50)	7.11 (2.16)
13.3. Incompatible beliefs	9	4.50 (3.87)	6.25 (1.71)	6.47 (1.92)
13.5. Identity associated with changed behavior	12	4.50 (3.00)	6.75 (0.50)	6.94 (2.44)
15.1. Verbal persuasion about capability	9	5.25 (0.50)	7.00 (0.82)	6.42 (2.06)
15.2. Mental rehearsal of successful performance	11	5.25 (0.96)	6.00 (0.82)	4.50 (2.82)
15.3. Focus on past success	8	6.75 (1.26)	7.50 (0.58)	7.00 (1.38)
15.4. Self-talk	5	5.25 (1.71)	8.50 (0.58)	7.70 (0.86)
A. Difficulties with side effects	8	8.25 (0.96)	7.75 (0.96)	7.06 (1.81)

BCT[a] or belief/concern	Number of messages	Relevance to improving medication adherence for people with type 2 diabetes, mean (SD)	Aim to have 8 to 10 messages that reflect the BCT or belief/concern well, mean (SD)	Fidelity of messages to the intended BCT/belief or concern[b], mean (SD)
B. Difficulties remembering and understanding the medication regimen (BCT 4.1: Instruction on how to perform a behavior)	8	7.50 (1.00)	7.50 (1.00)	7.22 (1.50)
C. Beliefs around medication in general and western medicines specifically	6	7.25 (0.50)	6.25 (0.96)	7.54 (1.28)
D. Perceived risks of taking medication (As E, BCT 5.1: Information about health consequences)	10	8.40 (0.89)	7.40 (0.55)	7.76 (1.02)
E. Beliefs about medication necessity (As D, BCT 5.1: Information about health consequences)	11	8.80 (1.30)	8.00 (0.71)	7.96 (1.07)
F. Social influence around taking medications	10	7.80 (0.84)	8.20 (0.45)	7.64 (1.24)
G. Health care system-related concerns	15	8.00 (0.71)	8.60 (0.89)	8.00 (1.01)

[a]BCT: behavior change technique.

[b]Mean is across reviewers per message and across messages per BCT.

Table 2. Reasons for exclusion of behavior change techniques following study 1.

Behavior change technique	Reason for exclusion
8.4. Habit reversal	Missing taking medication was not thought to be habit; therefore it was thought there was no habit to reverse in the sense psychologists think of habits
11.3. Conserving mental resources	Participants couldn't find a way to operationalize this that wasn't also habit formation. In addition, there was concern that suggesting people with type 2 diabetes focus on medication adherence alone could devalue other important lifestyle messages (eg, diet and physical activity) that are also key to diabetes self-management
12.2. Restructuring the social environment	Participants found it hard to structure messages that didn't seem to suggest stopping seeing people
13.1. Identification of self as a role model	Participants did not feel a single health behavior such as taking medication as prescribed constituted an identity and therefore would be hard to model oneself as a role model
13.2. Framing/reframing	Hard to use without interaction; would need to know how someone is framing to start with in order to initiate change. This was not thought to be possible in a text message
13.3. Incompatible beliefs	Thought to be better used in a therapeutic situation rather than via text message
15.3. Focus on past success	Designed to react to what someone has said, this would be difficult to deliver outside of a therapeutic situation as the text message system would not know if someone had had past successes or not

Study 3: Message Acceptability Survey

Participants

A total of 72 patients consented, of whom 11 rated no messages and 52 completed ratings for all messages. Patients who consented were aged between 30 and 83 (mean 62.0 [SD 11.3]) years, 39% (28/71; 1 participant did not specify) were female; 54% (39/72) of participants had had their last change in diabetes tablets over a year ago. See Table 3 for details and comparison with those who completed all the message reviews.

Table 3. Demographics and questionnaire responses for study 3 participants.

Characteristics	Consented (n=72)	Completed the message reviews (n=52)[a]
Female, n (%)[b]	28 (39)	21 (40)
Age in years, mean (SD)	61.99 (11.25)	62.81 (10.83)
Ethnicity, n (%)		
White	64 (89)	45 (87)
Asian/Asian British	1 (1)	0 (0)
Mixed/multiple ethnic groups	1 (1)	1 (2)
Did not answer	6 (8)	6 (12)
Time since diagnosis with diabetes in years, mean (SD)[c]	12.1 (5.38)	12.73 (5.76)
Time since last change to diabetes tablets, n (%)[d]		
≤1 month	6 (8)	3 (6)
>1 month to <3 months	6 (8)	6 (12)
≥3 months to <6 months	8 (11)	8 (15)
≥6 months to <1 year	9 (13)	6 (12)
>1 year	39 (54)	27 (52)
First prescription	3 (4)	1 (2)
Did not answer	1 (1)	1 (2)

[a]Includes postal response and 1 respondent who answered questions about all but the last message.

[b]n=71 consented, n=51 completed due to missing value.

[c]n=38 consented, n=26 completed; there was a problem with one of the 4 questionnaires, this question did not allow respondents to answer, some answered in the next question and these are included here.

[d]n=71 consented, n=52 completed due to missing value.

Behavior Change Techniques

Mean scores for BCTs were highest for ease of understanding (median 4.24, range 3.71 to 4.60), then liking (median 3.33, range 2.92 to 3.85), and lowest for perceived usefulness (median 2.88, range 2.49 to 3.41). Overall acceptability (median 3.49, range 3.10 to 3.89) was over the midpoint of the scale (see Table 4).

Decisions for Message Development

As none of the BCTs scored below the midpoint on the scale (midpoint=3) for acceptability, it was decided to retain all the BCTs in the library. The number of messages in the library therefore remained at 306, based on 24 BCTs and 7 beliefs and concerns (3 of these beliefs and concerns were based on an additional 2 BCTs).

Table 4. Ease of understanding, liking, usefulness, and acceptability scores per behavior change technique or belief or concern from study 3. All available message reviews from participants (n=61) were analysed.

BCT[a]/belief or concern	Ease of understanding, mean (SD)	Liking, mean (SD)	Usefulness, mean (SD)	Overall acceptability, mean (SD)
1.2. Problem solving	4.13 (0.16)	3.33 (0.26)	2.84 (0.17)	3.43 (0.17)
1.4. Action planning	4.24 (0.09)	3.34 (0.18)	2.99 (0.20)	3.52 (0.12)
2.3. Self-monitoring of behavior	4.48 (0.06)	3.45 (0.25)	2.76 (0.20)	3.56 (0.13)
3.1. Social support (unspecified)	4.30 (0.17)	3.41 (0.06)	2.95 (0.17)	3.55 (0.11)
3.2. Social support (practical)	4.25 (0.16)	3.20 (0.08)	2.68 (0.28)	3.38 (0.05)
3.3. Social support (emotional)	4.22 (0.15)	3.27 (0.05)	2.91 (0.24)	3.47 (0.04)
4.2. Information about antecedents	4.43 (0.10)	3.68 (0.22)	3.25 (0.14)	3.79 (0.14)
5.3. Information about social and environmental consequences	4.38 (0.25)	3.28 (0.17)	2.85 (0.27)	3.50 (0.15)
5.5. Anticipated regret	4.46 (0.22)	2.95 (0.20)	2.88 (0.70)	3.43 (0.37)
5.6. Information about emotional consequences	4.08 (0.14)	3.35 (0.21)	3.04 (0.45)	3.49 (0.26)
6.2. Social comparison	4.18 (0.13)	3.25 (0.38)	2.79 (0.15)	3.41 (0.22)
6.3. Information about others' approval	4.29 (0.15)	2.96 (0.43)	2.60 (0.42)	3.28 (0.24)
7.1. Prompts/cues	4.48 (0.10)	3.41 (0.26)	2.88 (0.30)	3.59 (0.18)
8.3. Habit formation	4.26 (0.21)	3.65 (0.10)	3.29 (0.40)	3.73 (0.13)
9.1. Credible source	4.53 (0.16)	3.56 (0.22)	3.11 (0.33)	3.73 (0.16)
9.2. Pros and cons	3.71 (0.56)	2.92 (0.42)	2.67 (0.31)	3.10 (0.41)
9.3. Comparative imaginings of future outcomes	3.92 (0.27)	3.12 (0.21)	2.93 (0.10)	3.32 (0.17)
10.5. Social reward	4.22 (0.53)	3.02 (0.39)	2.57 (0.43)	3.27 (0.45)
11.2. Reduce negative emotions	4.18 (0.06)	3.09 (0.27)	2.82 (0.14)	3.36 (0.10)
12.1. Restructuring the physical environment	4.60 (0.08)	3.62 (0.29)	3.09 (0.28)	3.77 (0.18)
13.5. Identity associated with changed behavior	4.00 (0.06)	3.20 (0.09)	2.93 (0.15)	3.38 (0.06)
15.1. Verbal persuasion about capability	4.22 (0.17)	3.32 (0.05)	2.73 (0.06)	3.43 (0.07)
15.2. Mental rehearsal of successful performance	3.87 (0.19)	2.95 (0.25)	2.57 (0.21)	3.13 (0.19)
15.4. Self-talk	4.35 (0.20)	3.25 (0.27)	2.88 (0.39)	3.49 (0.17)
A: Difficulties with side effects	4.40 (0.17)	3.85 (0.25)	3.41 (0.12)	3.89 (0.18)
B: Difficulties remembering and understanding the medication regimen (BCT 4.1: Instruction on how to perform a behavior)	4.11 (0.22)	3.33 (0.08)	2.83 (0.11)	3.42 (0.02)
C: Beliefs around medication in general and western medicines specifically	4.23 (0.10)	3.38 (0.16)	2.86 (0.06)	3.49 (0.09)
D: Perceived risks of taking medication (As E, BCT 5.1: Information about health consequences)	4.09 (0.56)	3.35 (0.55)	3.07 (0.45)	3.50 (0.52)
E: Beliefs about medication necessity (As D, BCT 5.1: Information about health consequences)	4.46 (0.12)	3.70 (0.05)	3.27 (0.09)	3.81 (0.03)
F: Social influence around taking medications (BCT 3.1: Social support unspecified)	4.35 (0.15)	3.23 (0.18)	2.49 (0.30)	3.36 (0.16)
G: Health care system–related concerns	4.04 (0.36)	3.48 (0.54)	3.10 (0.58)	3.54 (0.49)

[a]BCT: behavior change technique.

Study 4: Message Fidelity Survey

Participants

A total of 18 participants completed the survey; all identified themselves as researchers (rather than clinicians or researchers and clinicians). They had been paid to do research for between 6 and 31 (mean 14.6 [SD 6.0]) years and had published between 5 and 110 papers in the areas of diabetes, medication adherence, or behavioral interventions (mean 29.2 [SD 29.0]). Behavior change interventions were rated as being central or somewhat central to all 18 participants. Medication adherence was also reported as central or somewhat central to 13 participants, and diabetes as central or somewhat central to 7 participants.

Behavior Change Techniques

The mean scores of fidelity to the intended BCT ranged from 5.53 (BCT 4.1: Instruction on how to perform a behavior) to 8.87 (BCT: 5.3 Information about social and environmental consequences) with a mean of 7.61 (SD 0.93). See Table 5.

Decisions for Message Development

As none of the BCTs fell below the midpoint on the scale for fidelity (midpoint=5.5) it was decided that all the BCTs examined could be adequately represented by brief messages. At this stage of development, the findings related to individual messages from studies 2, 3, and 4 were considered together. One message was removed as it had scored below the midpoint for acceptability to patients in study 3 and for fidelity to the intended BCT in study 4 ("Diabetes tablets do not wear off over time" intended to represent BCT 5.1: Information about health consequences).

The final stage of message development was to review each message from the library with HCPs working on the project to ensure there was nothing either misleading or that could contradict current medical advice. One BCT (10.5: Social reward) was removed as the messages relied on a link to pharmacy pick-up that was not possible from the message delivery system at this time. A further 3 messages were removed and 7 were amended as a result of this review (see Table 6 for reasons). The final library contained 290 messages based on 25 BCTs and 4 beliefs and concerns (see Table B in Multimedia Appendix 1 for sample messages and Figure 1 for a flowchart of the message development process).

Table 5. Fidelity of the messages to the intended behavior change techniques from study 4.

Behavior change technique	Fidelity of messages mean (SD)
1.2. Problem solving	8.73 (0.54)
1.4. Action planning	7.54 (2.22)
2.3. Self-monitoring of behavior	7.79 (0.80)
3.1. Social support (unspecified)	7.91 (0.74)
3.2. Social support (practical)	8.73 (0.41)
3.3. Social support (emotional)	8.07 (0.20)
4.1. Instruction on how to perform a behavior	5.53 (0.93)
4.2. Information about antecedents	6.43 (0.62)
5.1. Information about health consequences	6.49 (1.97)
5.3. Information about social and environmental consequences	8.87 (0.19)
5.5. Anticipated regret	8.43 (1.16)
5.6. Information about emotional consequences	6.67 (1.69)
6.2. Social comparison	7.17 (1.45)
6.3. Information about others' approval	8.44 (0.92)
7.1. Prompts/cues	8.32 (0.45)
8.3. Habit formation	8.32 (1.06)
9.1. Credible source	7.49 (1.71)
9.2. Pros and cons	8.46 (0.65)
9.3. Comparative imaginings of future outcomes	7.02 (1.81)
10.5. Social reward	5.76 (0.55)
11.2. Reduce negative emotions	6.73 (0.49)
12.1. Restructuring the physical environment	8.40 (0.51)
13.5. Identity associated with changed behavior	7.03 (1.18)
15.1. Verbal persuasion about capability	7.75 (1.64)
15.2. Mental rehearsal of successful performance	7.48 (0.97)
15.4. Self-talk	8.13 (0.88)

Table 6. Reasons for message removal or amendment following health care provider review.

Message	BCT[a]	Action	Reason
If your blood sugar gets out of control, it could put you on an emotional rollercoaster. Taking your tablets as often as you should will stop the ups and downs.	5.6. Information about emotional consequences	Removed	Message implies a direct link between blood sugar control and an emotional rollercoaster—this is not true
Some people with diabetes who don't take their tablets lose their foot. Imagine how you feel if you let this happen to yourself.	5.5. Anticipated regret	Removed	Message implies a direct link between not taking tablets and losing a foot—smoking is a bigger risk factor.
Most religious leaders would agree that taking diabetes medication regularly is important.	9.1. Credible sources	Removed	Message flagged in qualitative comments of study 3; was not thought to be wholly accurate.
The tablets prescribed for your diabetes are very effective and can work alongside other treatments. Please let us know of anything else you may be taking.	C. Beliefs around medication in general and western medicines specifically	Amended	Confusing whom the message is seen as coming from.
If you don't understand don't be afraid to say so. There's a team of people here to help you.	G. Health care system-related concerns	Amended	Confusing whom the message is seen as coming from.
Taking your tablets can be as routine as having your morning coffee. Use this time as a prompt, and make your tablet taking a habit.	8.3. Habit formation	Amended	May have a morning coffee at 11:00 without any food—tablets should be taken with food.
Don't end up feeling guilty about extra diabetes complications caused by not taking your tablets as prescribed.	11.2. Reduce negative emotions	Amended	Complications may not be caused by not taking your tablets; referring to increased risk of complications is more accurate
If you lost your eyesight because you didn't take your diabetic tablets, would you regret it?	5.5. Anticipated regret	Amended	Loss is very final, seen to be too negative as this may go to people who have eyesight problems and these may not be linked to diabetic medication—a link implied that isn't necessarily there.
If you take your tablets as prescribed, it reduces your risk of serious complications by 50% (Diabetes UK).	9.1. Credible sources	Amended	Quantification hard to substantiate, so removed
Diabetes UK: Forgetting diabetes tablets just twice a week halves their overall benefit (hyperlink to webpage).	9.1. Credible sources	Amended	Quantification hard to substantiate, so removed

Discussion

Following this systematic development process involving four studies, we conclude that the messages produced have good acceptability and fidelity to their intended BCT. The retained BCTs were shown to be appropriate for delivery through brief messages, acceptable to patients, and clearly understood. All mean scores for retained BCTs were above the median point on the scales in studies 1, 3, and 4.

There are at least three key strengths of this research. First and most important, the systematic transparent process of message development allows for clarity in terms of message content; this will aid interpretation of any intervention effects. Second, the inclusion of people with type 2 diabetes, experts in behavior change research, and health care practitioners to ensure views of key stakeholders are incorporated. Third, as a consequence of the first two strengths, this research has resulted in the identification of BCTs appropriate for delivery through brief messages and development of a library of messages with evidence supporting their acceptability and fidelity to explicit BCTs that can be used in further research.

The majority of the sample of people with type 2 diabetes recruited for studies 2 and 3 were of white ethnicity, and this is acknowledged as a limitation. However, the work reported here is part of a larger program of work to develop evidence-based brief messaging for type 2 diabetes. Due to the increased risk of the condition among South Asian populations [22], these populations are a specific focus of this work. Focus groups with South Asian populations were run in parallel with those described in study 2 [23]. Further development will be needed to explore the similarities and differences between these two focus group studies and the potential tailoring, or additional messages needed, to ensure a library of messages that are acceptable to the whole population of people with type 2 diabetes.

As identified in a Cochrane review, novel approaches to medication adherence are needed [4]. We were therefore keen at this stage to include as many potentially effective components as possible for further testing in a trial setting while excluding any messages or BCTs that were definitely unacceptable to the target population. Currently there is not a widely recognized threshold for acceptability for these studies. A threshold for acceptability was set at the median point or above [24], as this seemed appropriate to meet our aims. In a recent definition of acceptability, cognitive and emotional reactions to an intervention and perceived relevance were identified as key facets in the definition [25]. In addition to formative assessment of the acceptability of the broad concepts used in study 2, these individual facets have been measured in study 3 through ease of understanding, liking, and perceived usefulness to an individual. Of the facets of acceptability measured, perceived

usefulness scored lower than either liking or ease of understanding. It should be noted that at this stage in development it was only possible to measure anticipated acceptability. As conceptualized by Sekhon and colleagues [25], acceptability should be measured across development of an intervention to assess both anticipated and experienced acceptability. It could be hypothesized that of the facets measured, perceived usefulness is the hardest to prospectively measure accurately. Future research could explore whether certain approaches may be perceived as more or less useful to different people in different contexts as this might be more appropriately done when people have had a chance to use the intervention in a more real-world setting and can report experienced acceptability. Future research could also explore the relationship between anticipated acceptability as we have measured here and experienced acceptability when participants are receiving these messages as part of their day-to-day lives.

Previous research in this area has found modest improvements in glycemic control in those with poorly controlled diabetes as a result of receiving brief diabetes-related messages in conjunction with blood glucose monitoring [26]. The messages in this previous research were generated within the research team and based on a review of previous research, existing mobile health interventions, and current patient resources [27]. The intervention was described as informed by 2 behavior change theories, and researchers incorporated 8 BCTs from an early version of the taxonomy [14] into the messages including "prompting self-monitoring" and "providing general encouragement" as well as reminders and feedback on blood glucose results through messages and an accompanying website [27]. It is not clear from the publications how the theories and BCTs were selected from the sources of information used (eg, the review, mobile health interventions, and patient resources). The messages were reviewed by diabetes experts and people with type 2 diabetes and piloted for acceptability and usability [27]. By contrast, the 4 studies reported in our research expand on this previous research by using a wider pool of individuals to generate the messages, looking beyond the individual theories and BCTs currently used to incorporate a wider range and back-checking with a separate group of researchers that the messages are good exemplars of the intended BCTs.

This approach may lead to more novelty in brief message interventions and greater confidence that individuals who receive the messages will be receiving the intended technique. This is important because if the intervention is found to be either effective or not effective, it will allow researchers to explain why, which will add to the empirical evidence in this area and allow researchers to optimize the intervention. Although some of the BCTs identified in the rapid review were deemed by behavior change researchers and health care professionals to be difficult to deliver through brief messages, we retained 29 BCTs and beliefs and concerns through to the final message library. This ensures that in addition to the more commonly used BCTs such as "self-monitoring of behavior," this intervention can explore approaches to this problem that are novel in brief message interventions.

This research provides a transparent development process for other researchers to use or improve upon. While this approach is not intended to be an alternative to other approaches of overall intervention development, to the authors' knowledge there is currently no guide for systematically developing the content of messages used in self-management interventions and ensuring that the content accurately represents the intended BCTs. This approach could therefore be used in conjunction with other approaches to intervention development [8]. The components of this approach were all important, however the order the studies were undertaken could have been altered. Initially, we planned to conduct the focus group study (reported here as study 2) prior to the message development study (reported here as study 1). As one of the aims of the focus group study was to ascertain people with diabetes' views on a system such as this, if these had been negative, the time and effort in developing the messages may have been wasted. However, due to delays in recruitment for the focus group study, we decided to go ahead with the message development study for two reasons: (1) the brief messages developed could be delivered in a number of ways (if participants had favored an app or Web-based system or even a paper-based system, the text of the messages could still have been used) and (2) it is easier for people to give opinions on a proposed system if they are provided with concrete examples. The opinions given by participants in study 2 were more useful as they were of actual messages we were planning to use rather than any that had been generated within the team. Overall, the process was relatively time consuming compared with other ways of generating messages; however, this is considered justified for the confidence we can now have that these messages are acceptable and have fidelity to the intended BCTs. Furthermore, there is the potential that some of this work could be adapted to other medication adherence interventions without repeating all stages; this may be less time consuming than developing those interventions from scratch. There were some decisions the team made about BCTs to include based on the available technology, which is constantly evolving. Because we have transparently presented the BCTs considered and reasons for exclusion, we hope if other researchers are considering other forms of technology or systems with a greater number of components, the excluded BCTs could duly be reconsidered.

Through developing a library of brief messages that are based on explicit BCTs and have been shown to be acceptable to a patient population, this research could also provide a valuable resource to develop theory in the area of medication adherence and diabetes and to optimize brief message interventions for this population. In a recently published review study, researchers explored how BCTs might have an effect on behavior by extracting the proposed links between BCTs and mechanisms of action [28]. However, the authors noted that the links proposed by intervention study authors were rarely tested empirically. Due to the fidelity of the messages to explicit BCTs, the messages developed here could be used to test theories of behavior change and the relationships between BCTs, mechanisms of action, and behavior change. The message library developed here will be further refined through our forthcoming pilot and feasibility work. Although messages were found to have good fidelity to their intended BCT, it was raised in qualitative comments that a single message may represent more than one BCT from within the taxonomy [15]; this could be

explored in future research. In addition, it was recognized that not all BCTs are suitable for delivery through brief messages. Specifically, those requiring interaction could not be facilitated by the current system. Further research could explore more advanced systems that could deliver these BCTs (eg, chat bots to allow lifelike automated responses during interactions), but it must be accepted that there may be some BCTs that are better delivered in person. In these cases, future research could explore the possibility of incorporating any of these techniques into medication reviews with an HCP.

Participants in the study 2 focus groups requested messages around diet and exercise management in addition to messages focused on medication adherence. A wariness was also expressed by the participants in study 1 that focusing solely on medication adherence may have the unintended effect of making other aspects of diabetes self-management seem less important. Messages that provide information from verified sources related to other aspects of diabetes care therefore may form an important part of engaging someone with the system and ensure a more holistic view of the condition. However, this will create a challenge when interpreting results and exploring the mechanisms of action. In addition, the focus here was on tablet medication adherence; further research would be needed to assess whether any of these messages would be acceptable and appropriate for people using only injected medication or what adaptations would be needed to ensure relevance. Our future work will aim to explore the potential mechanisms of action both qualitatively and quantitatively with participants while using the messaging system.

In conclusion, a library of brief messages acceptable to people with type 2 diabetes representing explicit BCTs has been developed using a rigorous, transparent process. This will provide the basis for a novel brief message intervention to improve medication adherence in people with type 2 diabetes and can be further used to develop the theory and understanding of behavior change in this area.

Acknowledgments

The authors would like to acknowledge the support of Katherine Grady and the Research for the Future/Help BEAT campaigns (National Institute for Health Research [NIHR] Clinical Research Network Greater Manchester), Northern Care Alliance NHS Group. We would also like to acknowledge the experts in behavior change research and HCPs who contributed to this work: Madelynne Arden, Leah Avery, Ben Brown, Jamie Brown, Sarah Chapman, Angel Chater, Anna Chisolm, Neil Coulson, Sinead Currie, Stephan Dombrowski, Frank Doyle, Ben Gardner, Kyra Hamilton, David Hevey, Shashi Hirani, Gareth Hollands, Delyth James, Ian Kellar, Keegan Knittle, Sheila McCorkindale, Rosie McEachan, Jenny McSharry, Gerry Molloy, Leanne Morrison, Val Morrison, Felix Naughton, Katie Newby, Nicki O'Brien, Ronan O'Carroll, Andrew Prestwich, Urte Scholz, Debbie Smith, Jackie Steadman, Sarah Tonkin-Crine, Lisa Marie Warner, Thomas Webb, Stefanie Williams, Caroline Wood, and Alison Wright, as well as Tracy Epton, Hannah Long and Laura Brown, the colleagues who helped to facilitate the study 1 workshop; Nikki Newhouse, who helped with the message library review post study 3; Suman Prinjha, who designed and facilitated focus groups with South Asian populations; and the wider Support through Mobile Messaging and digital health Technology for Diabetes [SuMMiT-D] team. This publication presents independent research funded by the NIHR under its Programme Grants for Applied Research program (RP-PG-1214-20003) and supported by the NIHR Oxford Biomedical Research Centre. AF is an NIHR Senior Investigator. The views expressed are those of the authors and not necessarily those of NIHR or the Department of Health and Social Care. The SuMMiT-D research team acknowledges the support of the NIHR Clinical Research Network.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables.
[DOCX File , 16 KB - jmir_v22i5e15989_app1.docx]

References

1. National Health Service. 2017. What is type 2 diabetes? URL: https://www.nhs.uk/conditions/type-2-diabetes/ [accessed 2020-02-26]
2. Polonsky WH, Henry RR. Poor medication adherence in type 2 diabetes: recognizing the scope of the problem and its key contributors. Patient Prefer Adherence 2016;10:1299-1307 [FREE Full text] [doi: 10.2147/PPA.S106821] [Medline: 27524885]
3. Farmer AJ, Rodgers LR, Lonergan M, Shields B, Weedon MN, Donnelly L, et al. Adherence to oral glucose-lowering therapies and associations with 1-year HbA1c: a retrospective cohort analysis in a large primary care database. Diabetes Care 2016 Feb;39(2):258-263 [FREE Full text] [doi: 10.2337/dc15-1194] [Medline: 26681714]
4. Nieuwlaat R, Wilczynski N, Navarro T, Hobson N, Jeffery R, Keepanasseril A, et al. Interventions for enhancing medication adherence. Cochrane Database Syst Rev 2014;11:CD000011. [doi: 10.1002/14651858.CD000011.pub4] [Medline: 25412402]

5. Farmer AJ, McSharry J, Rowbotham S, McGowan L, Ricci-Cabello I, French DP. Effects of interventions promoting monitoring of medication use and brief messaging on medication adherence for people with Type 2 diabetes: a systematic review of randomized trials. Diabet Med 2015 Oct 15;33(5):565-579. [doi: 10.1111/dme.12987] [Medline: 26470750]
6. Ybarra ML, Holtrop JS, Bağci BAT, Emri S. Design considerations in developing a text messaging program aimed at smoking cessation. J Med Internet Res 2012;14(4):e103 [FREE Full text] [doi: 10.2196/jmir.2061] [Medline: 22832182]
7. Maar MA, Yeates K, Toth Z, Barron M, Boesch L, Hua-Stewart D, et al. Unpacking the black box: a formative research approach to the development of theory-driven, evidence-based, and culturally safe text messages in mobile health interventions. JMIR Mhealth Uhealth 2016 Jan 22;4(1):e10 [FREE Full text] [doi: 10.2196/mhealth.4994] [Medline: 26800712]
8. O'Cathain A, Croot L, Sworn K, Duncan E, Rousseau N, Turner K, et al. Taxonomy of approaches to developing interventions to improve health: a systematic methods overview. Pilot Feasibility Stud 2019;5:41 [FREE Full text] [doi: 10.1186/s40814-019-0425-6] [Medline: 30923626]
9. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
10. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
11. Crutzen R. The behavioral intervention technology model and intervention mapping: the best of both worlds. J Med Internet Res 2014 Aug 05;16(8):e188 [FREE Full text] [doi: 10.2196/jmir.3620] [Medline: 25095730]
12. Whittaker R, Merry S, Dorey E, Maddison R. A development and evaluation process for mHealth interventions: examples from New Zealand. J Health Commun 2012;17 Suppl 1:11-21. [doi: 10.1080/10810730.2011.649103] [Medline: 22548594]
13. Abroms LC, Whittaker R, Free C, Mendel VAJ, Schindler-Ruwisch JM. Developing and pretesting a text messaging program for health behavior change: recommended steps. JMIR Mhealth Uhealth 2015 Dec 21;3(4):e107 [FREE Full text] [doi: 10.2196/mhealth.4917] [Medline: 26690917]
14. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]
15. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
16. Dobson R, Whittaker R, Pfaeffli Dale L, Maddison R. The effectiveness of text message-based self-management interventions for poorly-controlled diabetes: A systematic review. Digit Health 2017;3:2055207617740315 [FREE Full text] [doi: 10.1177/2055207617740315] [Medline: 29942620]
17. Kebede MM, Liedtke TP, Möllers T, Pischke CR. Characterizing active ingredients of ehealth interventions targeting persons with poorly controlled type 2 diabetes mellitus using the behavior change techniques taxonomy: scoping review. J Med Internet Res 2017 Oct 12;19(10):e348 [FREE Full text] [doi: 10.2196/jmir.7135] [Medline: 29025693]
18. Long H, Bartlett YK, Farmer AJ, French DP. Identifying brief message content for interventions delivered via mobile devices to improve medication adherence in people with type 2 diabetes mellitus: a rapid systematic review. J Med Internet Res 2019 Jan 09;21(1):e10421 [FREE Full text] [doi: 10.2196/10421] [Medline: 30626562]
19. National Institute of Health Research Involve. Policy on payments and expenses for members of the public 2016 URL: https://www.invo.org.uk/resource-centre/payment-and-recognition-for-public-involvement/ [accessed 2019-08-01]
20. Bartlett YK, Newhouse N, Long HA, Farmer A, French D. What do people with type 2 diabetes want from a brief messaging system to support medication adherence? Patient Prefer Adherence 2019;13:1629-1640 [FREE Full text] [doi: 10.2147/PPA.S217843] [Medline: 31686789]
21. Curtis KE, Lahiri S, Brown KE. Targeting parents for childhood weight management: development of a theory-driven and user-centered healthy eating app. JMIR Mhealth Uhealth 2015;3(2):e69 [FREE Full text] [doi: 10.2196/mhealth.3857] [Medline: 26088692]
22. Gujral UP, Pradeepa R, Weber MB, Narayan KMV, Mohan V. Type 2 diabetes in South Asians: similarities and differences with white Caucasian and other populations. Ann N Y Acad Sci 2013 Apr;1281:51-63 [FREE Full text] [doi: 10.1111/j.1749-6632.2012.06838.x] [Medline: 23317344]
23. Prinjha S, Ricci-Cabello I, Newhouse N, Farmer A. British South Asian Patients' Perspectives on the Relevance and Acceptability of Mobile Health Text Messaging to Support Medication Adherence for Type 2 Diabetes: Qualitative Study. JMIR Mhealth Uhealth 2020 Apr 20;8(4):e15789 [FREE Full text] [doi: 10.2196/15789] [Medline: 32310150]
24. Naughton F, Cooper S, Foster K, Emery J, Leonardi-Bee J, Sutton S, et al. Large multi-centre pilot randomized controlled trial testing a low-cost, tailored, self-help smoking cessation text message intervention for pregnant smokers (MiQuit). Addiction 2017 Jul;112(7):1238-1249 [FREE Full text] [doi: 10.1111/add.13802] [Medline: 28239919]
25. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res 2017 Jan 26;17(1):88 [FREE Full text] [doi: 10.1186/s12913-017-2031-8] [Medline: 28126032]

26. Dobson R, Whittaker R, Jiang Y, Shepherd M, Maddison R, Carter K, et al. Text message-based diabetes self-management support (SMS4BG): study protocol for a randomised controlled trial. Trials 2016 Apr 02;17:179 [FREE Full text] [doi: 10.1186/s13063-016-1305-5] [Medline: 27039300]
27. Dobson R, Carter K, Cutfield R, Hulme A, Hulme R, McNamara C, et al. Diabetes text-message self-management support program (SMS4BG): a pilot study. JMIR Mhealth Uhealth 2015;3(1):e32 [FREE Full text] [doi: 10.2196/mhealth.3988] [Medline: 25830952]
28. Carey RN, Connell LE, Johnston M, Rothman AJ, de Bruin M, Kelly MP, et al. Behavior change techniques and their mechanisms of action: a synthesis of links described in published intervention literature. Ann Behav Med 2019 Jul 17;53(8):693-707 [FREE Full text] [doi: 10.1093/abm/kay078] [Medline: 30304386]

Abbreviations

BCT: behavior change technique
HCP: health care professional
NHS: National Health Service
NIHR: National Institute for Health Research
PPI: Patient and Public Involvement panel
SMS: short message service

Edited by G Eysenbach; submitted 25.08.19; peer-reviewed by K Newby, G Molloy, M Jones; comments to author 08.10.19; revised version received 25.10.19; accepted 16.12.19; published 13.05.20.

Please cite as:
Bartlett YK, Farmer A, Rea R, French DP
Use of Brief Messages Based on Behavior Change Techniques to Encourage Medication Adherence in People With Type 2 Diabetes: Developmental Studies
J Med Internet Res 2020;22(5):e15989
URL: https://www.jmir.org/2020/5/e15989
doi:10.2196/15989
PMID:32401214

Original Paper

Gratitude at Work: Prospective Cohort Study of a Web-Based, Single-Exposure Well-Being Intervention for Health Care Workers

Kathryn C Adair[1], PhD; Larissa G Rodriguez-Homs[2], MBS; Sabran Masoud[2], BS; Paul J Mosca[3,4], MA, PhD; J Bryan Sexton[1,5], PhD

[1]Duke Center for Healthcare Safety and Quality, Duke University Health System, Durham, NC, United States

[2]Duke University School of Medicine, Durham, NC, United States

[3]Duke Network Services Duke University Health System, Duke University Health System, Durham, NC, United States

[4]Department of Surgery, Duke University School of Medicine, Duke University Health System, Durham, NC, United States

[5]Department of Psychiatry, Duke University School of Medicine, Duke University Health System, Durham, NC, United States

Corresponding Author:
Kathryn C Adair, PhD
Duke Center for Healthcare Safety and Quality
Duke University Health System
University Tower, Suite 300
3100 Tower Blvd
Durham, NC, 27707
United States
Phone: 1 9196686812
Fax: 1 9196680484
Email: kathryn.c.adair@duke.edu

Abstract

Background: Emotional exhaustion (EE) in health care workers is common and consequentially linked to lower quality of care. Effective interventions to address EE are urgently needed.

Objective: This randomized single-exposure trial examined the efficacy of a gratitude letter–writing intervention for improving health care workers' well-being.

Methods: A total of 1575 health care workers were randomly assigned to one of two gratitude letter–writing prompts (self- vs other focused) to assess differential efficacy. Assessments of EE, subjective happiness, work-life balance, and tool engagement were collected at baseline and 1-week post intervention. Participants received their EE score at baseline and quartile benchmarking scores. Paired-samples *t* tests, independent *t* tests, and correlations explored the efficacy of the intervention. Linguistic Inquiry and Word Count software assessed the linguistic content of the gratitude letters and associations with well-being.

Results: Participants in both conditions showed significant improvements in EE, happiness, and work-life balance between the intervention and 1-week follow-up ($P<.001$). The self-focused (vs other) instruction conditions did not differentially predict improvement in any of the measures ($P=.91$). Tool engagement was high, and participants reporting higher motivation to improve their EE had higher EE at baseline ($P<.001$) and were more likely to improve EE a week later ($P=.03$). Linguistic analyses revealed that participants high on EE at baseline used more negative emotion words in their letters ($P=.005$). Reduction in EE at the 1-week follow-up was predicted at the level of a trend by using fewer first-person ($P=.06$) and positive emotion words ($P=.09$). No baseline differences were found between those who completed the follow-up assessment and those who did not ($Ps>.05$).

Conclusions: This single-exposure gratitude letter–writing intervention appears to be a promising low-cost, brief, and meaningful tool to improve the well-being of health care workers.

(J Med Internet Res 2020;22(5):e15562) doi:10.2196/15562

KEYWORDS

burnout; health care; positive psychology; mental health; emotions

Introduction

Background

Emotional exhaustion (EE) among US health care workers has increased to 1 out of 3 nurses [1] and almost half of practicing physicians [2], with the highest rates among junior doctors and those working at the front line of patient care [3,4]. Many aspects of patient safety and quality of care appear compromised by EE, which is a key component of burnout [5]. EE has been linked to higher frequencies of medical errors, lapses in professionalism, impeded learning, and suicidal ideation [6-8]. As such, there is an urgent need for simple, brief, effective, and accessible EE and burnout interventions on a broad scale.

Unfortunately, existing remedies for EE are limited by time, effort, and cost. For instance, mindfulness-based meditation courses have empirical support for reducing EE [9]; however, courses of this type can be costly and typically require 8 to 10 weeks and approximately 90 hours of effort from health care workers who already skip meals and breaks, deprive themselves of needed rest, and get home late as a norm [10,11]. In contrast, this study examines the efficacy of a brief, single-use tool—writing a letter of gratitude—for reductions in EE and improvements in well-being. If there is a way to use bite-sized tools to provide a quick recharge in the face of EE, it could facilitate (not replace) the use of other, more elaborate remedies.

Gratitude has one of the strongest associations with better mental health and well-being of any personality trait, even more than hope, optimism, or compassion [12]. Several rigorous controlled trials demonstrate the beneficial effect of gratitude interventions [13-15]. For instance, participants randomized to gratitude journal keeping, compared with active control conditions, reported higher levels of energy, determination, enthusiasm, feeling more optimistic, being more likely to exercise, having fewer physical symptoms, sleeping longer, and sleeping with better quality [16]. Another study found that participants randomized to write a gratitude letter and visit the recipient of the letter reported significant gains in happiness and reductions in depressive symptoms, compared with a placebo control group, both 1 week and 1 month later [17]. In fact, this intervention exhibited the greatest benefits compared with four other positive psychology interventions that the authors tested (ie, 3 good things, you at your best, using signature strengths, and identifying signature strengths) [17].

Previous research on live interactions has shown that expressions of gratitude that focused more on the benefits that the person expressing the gratitude received (ie, self-focused) benefited the expresser, whereas gratitude focusing more on the positive attributes of the individual being thanked (ie, other focused) benefited the recipient of said gratitude [18]. We hypothesized that participants randomized to gratitude letter–writing instructions to focus more on personal benefits in their letters (ie, self-focus) would exhibit greater improvement in well-being than those randomized to focus more on the positive aspects of the letter recipient (ie, other-focused).

On the basis of previous linguistic studies that have identified a *language of depression*, we were interested in identifying a *language of burnout* within the gratitude letters [19]. As such, we developed a series of hypotheses regarding the frequencies of words in particular word categories and their associations with both concerning levels of baseline EE and EE improvement. More frequent use of first-person singular (eg, *I*) and negative emotion words and marginally fewer positive emotion words have been identified in those experiencing depression [19,20]. Greater use of cognitive processing words (eg, *because* and *realization*) predicts fewer posttraumatic stress disorder symptoms among trauma survivors [21], and in our own research, third-person plural use (eg, *we*) is higher in groups with lower EE [22]. Taking this research together, we expected participants with concerning levels of EE at baseline to use more first-person singular and negative emotion words and fewer third-person plural, positive emotion, and cognitive processing words in their letters. We expected that EE improvers would exhibit the inverse of this pattern.

Objectives

This trial examined the efficacy of a gratitude letter–writing intervention to improve health care workers' well-being measured in terms of EE, happiness, and work-life balance. The aims of this study were (1) to determine whether a onetime gratitude letter predicts improvements in well-being and to test whether self- vs other-focused instruction conditions differentially predict changes in well-being, (2) to measure participants' engagement with the intervention and their reactions to receiving feedback on their individual EE, and (3) to examine the linguistic content of gratitude letters for correlates of baseline EE and predictors of EE improvement.

Methods

Design and Patient Population

This randomized before-and-after trial of a gratitude letter–writing intervention was conducted between January 2018 and February 2019 (institutional review board approval: Pro00063703). Health care workers enrolled through a website link [23], having learned of the study through colleagues or by participating in a workshop conducted by authors KA and JS. All health care workers (clinical and nonclinical) aged at least 18 years were eligible to participate. Participants completed assessments and the gratitude letter–writing intervention at baseline and were given the opportunity to complete assessments again 1 week later. The intervention and assessments were completed on the web.

Among the 1575 health care workers who participated in the gratitude letter–writing intervention, 1179 (74.86%) wrote 15 words or more, a threshold at which we found letters were less likely to be left unfinished. A total of 17.59% (227/1575) participants completed both the pre- and postintervention assessments.

Measures

A total of 3 validated scales were used to assess aspects of well-being: EE, subjective happiness, and work-life balance. These scales were selected for the following reasons: (1) all are brief (8 items or less) and therefore are less onerous to busy health care workers to complete, (2) all are psychometrically

sound both in prior studies and in our own, and (3) all are sensitive to intervention in our prior studies [24]. Demographic questions on gender, race, ethnicity, and role were included at baseline.

Emotional Exhaustion

The Maslach Burnout Inventory (MBI) is widely used among professionals in human services, including health care workers, and is considered the gold standard survey instrument for assessing professional burnout [25]. The MBI includes 3 scales: EE, depersonalization, and personal accomplishment. EE consistently produces the largest and most consistent coefficient alpha estimates compared with depersonalization and personal accomplishment [5,26,27]. EE has also been used to discriminate between burned-out and nonburned-out outpatients suffering from work-related neurasthenia (according to International Classification of Diseases-10 criteria and Diagnostic and Statistical Manual of Mental Disorders-IV) [27,28]. We used a 5-item derivative of the original 9-item EE scale [25,29]. An example item is "I feel frustrated by my job." Participants respond using a 5-point scale (1=disagree strongly and 5=agree strongly), and mean scores are rescaled to 0 to 100, with higher scores signifying more EE. Internal consistency in this study was good (Cronbach alpha at baseline=.84 and Cronbach alpha at follow-up=.86).

To examine linguistic differences by level of EE, we compared the language of a priori selected word categories for those participants whose EE scores reflected moderate and severe EE (*concerning*) with participants without EE. An EE score greater than or equal to 75 was deemed concerning, which reflects, on average, agreeing slightly or strongly to EE items. The concerning threshold should not be considered clinically diagnostic, but rather, it identifies those whose scores suggest a pronounced level of EE [29]. Those determined as not having EE (not concerning) had scores less than 50, which reflects, on average, disagreeing slightly or strongly to EE items. Participants who scored between 50 and 75 (mild EE) were not included in these analyses. To compare linguistic markers of EE improvement, we identified *improvers* as having a reduction of their burnout score of more than 10 points and *decliners* as having scores that increased by 10 or more points.

Subjective Happiness

Lyubomirsky and Lepper's Subjective Happiness Scale (SHS) is a 4-item internationally used scale of global happiness [30,31]. An example item is "In general I consider myself (1=not a very happy person to 7=a very happy person)". SHS items are answered using a 7-point scale, and each participant's responses are averaged, with higher scores indicating higher happiness. Internal consistency in this study was good (Cronbach alpha at baseline=.83 and Cronbach alpha at follow-up=.87).

Work-Life Balance

The work-life climate scale evaluates individual differences in work-life infraction behaviors and has been shown to have good psychometrics when administered to health care workers [10,11,29,32]. The scale prompt asks *During the past week, how often did this occur*? Items include: *Arrived home late from work; Slept less than 5 hours in a night.* Responses include rarely or none of the time (less than 1 day), some or a little of the time (1-2 days), occasionally or a moderate amount of time (3-4 days), all of the time (5-7 days), and not applicable. Scale scores were computed by taking the mean of the items [10]. Internal consistency in this study was acceptable (Cronbach alpha at baseline=.71 and Cronbach alpha at follow-up=.77).

Intervention Experience

In addition, five questions assessed participants' experience with the tool. In the baseline survey, these questions were "Would you like a copy of this letter emailed to you?" (yes, no, or do not know) and "I completed this gratitude letter exercise before" (yes or no). The 1-week follow-up surveys included "Do you have other people in mind for whom you might write a letter like this?" (yes, no, or not sure), "Did you talk to anyone about your first gratitude letter"? (yes, no, or not sure), and "Since my first gratitude letter a week ago, it has gotten easier to think of things for which I am grateful" (1=very strongly agree and 8=very strongly disagree).

Emotional Exhaustion Feedback

After completing the items for the EE scale, participants were shown their scores and the quartiles of scores from our sample of over 135,000 US health care workers. Participants answered three Likert-like questions after receiving their scores and before the gratitude letter–writing intervention: "I was surprised by my burnout score," "Knowing my burnout score makes me want to work on it more," and "I would like to be more resilient." Response options ranged from 1 to 5 (disagree strongly to agree strongly).

Gratitude Intervention

The gratitude letter–writing intervention invited participants to spend approximately 7 min writing an appreciative letter to someone who has positively affected their lives.

Participants were given general instructions plus two text boxes (parts 1 and 2) to enter their letters. Participants were randomized, through an automatic randomizer within the survey software program, to receive either the *self-focus* condition or the *other-focus* condition as described below:

> *Think of someone who has done something amazing for you; this person can be alive or no longer with us. This person contributed to your well-being in a big way. Spend the next 7 minutes writing a genuine, kind and appreciative 2-part note:*

Self-focus condition:

> *Part 1:Tell this person what they did, how it impacted you, and the benefits you received.*
>
> *Part 2: Tell this person why it was important to you.*

Other-focus condition:

> *Part 1: Tell this person what they did, how it impacted you, how it made you feel, and why it was important to you.*
>
> *Part 2: Tell this person what it says about them, that they did this amazing thing for you. You might include what this says about your relationship to this person.*

Statistical Analysis

Demographic variables and evaluation questions were subjected to descriptive analysis. Cronbach alpha assessed the internal reliability of administered scales, with values above .70 considered acceptable [33]. Paired *t* tests assessed improvements in well-being between baseline and 1-week postintervention surveys. Independent-samples *t* tests compared changes in well-being scores between baseline and 1-week follow-up between gratitude letter–writing conditions. A manipulation check compared the linguistic use of first-person singular (eg, *I*) vs second person (eg, *you*) based on condition; the self-focus condition was expected to use more first-person words and fewer second-person words than the other-focus condition. Independent-samples *t* tests also compared EE feedback scores based on EE level at baseline (concerning vs no EE), EE improvement (improvers vs decliners), and baseline well-being scores in those who completed 1-week postintervention surveys vs those who did not. *P* values less than .05 were considered significant, and all tests were 2 tailed. Listwise deletion was used for missing data. Analyses were performed using SPSS version 24 software.

Linguistic Analysis of Letters

The Linguistic Inquiry and Word Count (LIWC) software program version 2015 was used to analyze text written in the participants' gratitude letters [34]. The LIWC dictionary consists of 6400 words, word stems, and select emotions, with each entry defining one or more of 80 language categories [35]. Word counts are expressed as a percentage of the total number of words, controlling for the length of the text file. Text from both parts 1 and 2 of the instructions were included in the linguistic analyses.

Independent-samples *t* tests compared differences in linguistic categories based on EE level at baseline (concerning vs no EE) and also compared differences in linguistic word count frequencies for EE improvers and decliners.

Results

Respondent Demographics

Overall, 1575 participants completed the onetime gratitude intervention and baseline assessments. The majority of participants were white and female (1165/1575, 73.84% and 1100/1575, 69.84%, respectively). The top 3 health care worker roles were those who classified as *other* (334/1575, 21.21%), *attending or staff physician* (266/1575, 16.89%), and *other manager* (260/1575, 16.51%). Nurse managers and charge nurses, when combined, accounted for 21.71% (342/1575). See Table 1 for additional demographic information.

Overall, 277 participants (self-focused condition=139 and other-focused condition=138) provided correct contact information and followed up at 1 week using a onetime email prompt. An independent *t* test revealed no differences in baseline EE, subjective happiness, or work-life balance for those who completed the 1-week follow-up and those who did not (*P*=.48, *P*=.44, and *P*=.12, respectively). At baseline, 38.02% (579/1523) participants met or exceeded the threshold for having concerning levels of EE.

Table 1. Respondent demographics of gratitude letter–writing participants (N=1575).

Demographics	Participants, n (%)
Health care worker role	
Other	334 (21.21)
Attending or staff physician	266 (16.89)
Fellow physician	6 (0.38)
Resident physician	6 (0.38)
Physician assistant, nurse practitioner, clinical nurse specialist	58 (3.68)
Nurse manager or charge nurse	175 (11.11)
Registered nurse (including certified registered nurse anesthetists)	167 (10.60)
Pharmacist	138 (8.76)
Therapist (respiratory therapy, physical therapy, occupational therapy, speech therapy)	5 (0.32)
Clinical social worker	15 (0.95)
Dietician or nutritionist	5 (0.32)
Clinical support (certified medical assistant, emergency medical technician, nurse aide, etc)	12 (0.76)
Technologist	8 (0.51)
Technician (eg, surgery, laboratory, radiology)	11 (0.70)
Admin support (clerk, secretary, receptionist)	23 (1.46)
Other manager (eg, clinic manager)	260 (16.51)
Missing	86 (5.46)
Gender	
Male	398 (25.30)
Female	1100 (69.84)
Missing	77 (4.89)
Race and ethnicity	
American Indian or Alaska Native	3 (0.19)
Asian	72 (4.57)
Black or African American	116 (7.40)
Hispanic or Latino	88 (5.6)
Native Hawaiian or other Pacific Islander	4 (0.25)
White	1165 (73.84)
Other	40 (2.54)
Missing	87 (5.52)

Aim 1: Assessing Change in Well-Being Metrics Between Baseline and 1-Week Follow-Up and Differential Efficacy by Instruction Conditions

Table 2 shows the means and *t* test results across the sample of participants who completed the 1-week follow-up. Examined separately, both conditions showed significant reductions in EE and improvements in happiness and work-life balance (*P*s<.001).

Table 2. Independent samples *t* tests comparing changes in well-being for participants in the self- vs other conditions.

Variable	EE[a] change: self (N=134), mean (SD)	EE change: other (N=135), mean (SD)	*t (df)*	*P* value	95% CI	Happiness change: self (N=136), mean (SD)	Happiness change: other (N=137), mean (SD)	*t (df)*	*P* value	95% CI	Work-life balance change: self (N=132), mean (SD)	Work-life balance change: other (N=134), mean (SD)	*t (df)*	*P* value	95% CI
Self- vs other focus	−7.98 (17.30)	−7.81 (16.85)	−0.08 (257)	.94	−4.27 to 3.93	5.26 (11.01)	4.93 (11.66)	0.24 (471)	.81	−2.37 to 3.04	−0.26 (0.45)	−0.26 (0.42)	−0.12 (264)	.91	−0.11 to 0.10

[a]EE: emotional exhaustion.

Table 3. Paired-samples *t* tests assessing well-being across participants who completed the self-focus condition, other-focus condition, and the overall sample at baseline and the 1-week follow-up.

Variable	EE[a] pre intervention, mean (SD)	EE post intervention, mean (SD)	*t (df)*	*P* value	95% CI	Happiness pre intervention, mean (SD)	Happiness post intervention, mean (SD)	*t (df)*	*P* value	95% CI	Work-life balance pre intervention, mean (SD)	Work-life balance post intervention, mean (SD)	*t (df)*	*P* value	95% CI
Self-focus	61.48 (25.56)[b]	53.51 (27.0)[b]	5.34 (133)	<.001	5.02 to 10.93	67.83 (19.7)[c]	73.09 (18.0)[c]	−5.57 (135)	<.001	−7.12 to −3.39	2.29 (0.61)[d]	2.03 (0.57)[d]	6.90 (131)	<.001	0.19 to 0.34
Other focus	59.88 (26.54)[e]	52.07 (28.48)[e]	5.38 (134)	<.001	4.93 to 10.67	66.84 (19.9)[f]	71.77 (17.4)[f]	−4.95 (136)	<.001	−6.90 to −2.96	2.28 (0.60)[b]	2.02 (0.59)[b]	7.27 (133)	<.001	0.18 to 0.32
Overall 1-week sample	60.67 (26.03)[g]	52.79 (27.7)[g]	7.59 (268)	<.001	5.84 to 9.94	67.34 (19.7)[h]	72.43 (17.7)[h]	−7.43 (272)	<.001	−6.44 to −3.74	2.28 (0.60)[i]	2.02 (0.58)[i]	10.02 (265)	<.001	0.21 to 0.32

[a]EE: emotional exhaustion.

[b]N=134.

[c]N=136.

[d]N=132.

[e]N=135.

[f]N=137.

[g]N=269.

[h]N=273.

[i]N=266.

When directly comparing change in well-being between self- and other-focused conditions, independent *t* tests revealed no significant differences across any of the three well-being measures (*P*s>.05; see Table 2). A linguistic manipulation check revealed that those randomized to the self-focused prompt used more first-person singular words (ie, *I*; self: mean 11.32, SD 4.021; other: mean 10.08, SD 3.66; t_{1496}=6.06; *P*<.001) and fewer second-person words (eg, *you*; self: mean 4.61, SD 4.04; other: mean 5.68, SD 4.47; t_{1496}=−4.85; *P*<.001) than those randomized to the other-focus condition, indicating that condition prompts did elicit differential linguistic self- vs other focus.

As conditions did not differ in their efficacy, we reran the paired *t* tests with the pooled sample to obtain the overall effects, which remained significant (*P*s<.001; see Table 3 and Figure 1). Participants with higher levels of EE at baseline were more likely to be improvers (mean 63.99, SD 23.17) compared with decliners (mean 50.21, SD 26.52) at the 1-week follow-up (t_{235}=3.89; *P*<.001). In addition, improvements in specific work-life balance behaviors were examined across the whole sample. Significant improvements were reported across all behaviors (see Table 4), with the biggest improvement in feeling *less* frustration with technology (t_{257}=6.18; *P*<.001).

Figure 1. Pre- and postintervention well-being scores. For demonstration purposes, the work-life balance variable was scaled 0-100 in this graph, original response range 1-4. Error bars represent SE means.

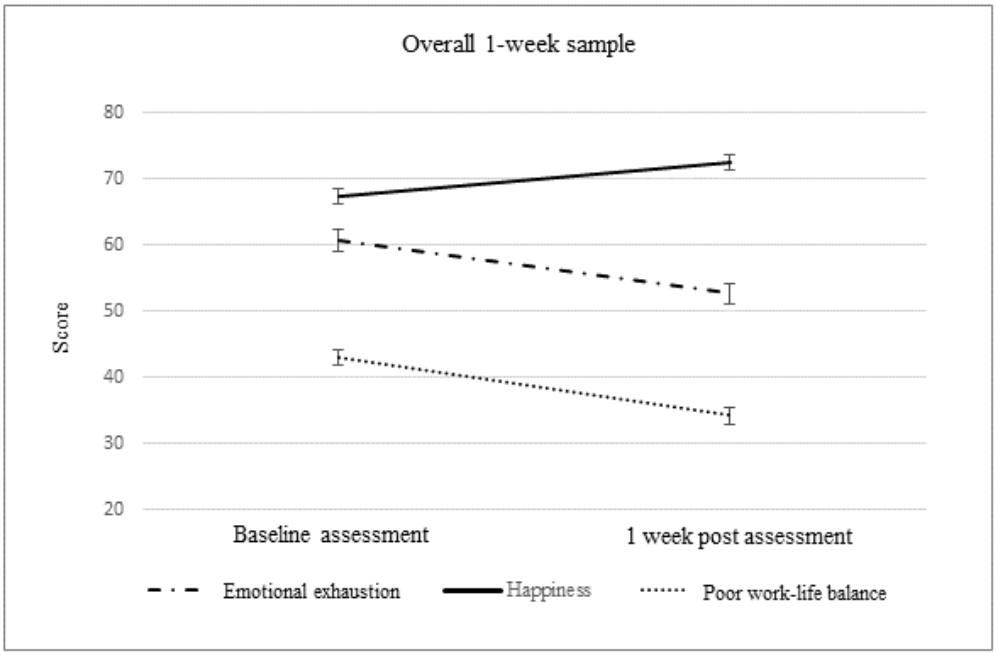


Table 4. Paired-samples *t* tests assessing work-life balance items across participants who completed the 1-week follow-up gratitude letter–writing intervention.

Variable	Baseline, mean (SD)	1-week follow-up			
		Mean (SD)	*t (df)*	*P* value	95% CI
Slept less than 5 hours in a night	1.91 (0.94)	1.76 (0.92)	2.68 (260)	.01	0.04 to 0.25
Worked through a shift or day without any breaks	2.32 (1.08)	2.09 (1.04)	4.23 (256)	<.001	0.13 to 0.35
Skipped a meal	2.13 (1.00)	1.88 (0.98)	4.49 (262)	<.001	0.14 to 0.36
Ate a poorly balanced meal	2.63 (0.93)	2.36 (0.99)	5.44 (263)	<.001	0.17 to 0.37
Had difficulty sleeping	2.40 (0.99)	2.10 (1.00)	5.36 (262)	<.001	0.19 to 0.42
Arrived home late from work	2.56 (1.03)	2.24 (0.96)	5.95 (256)	<.001	0.22 to 0.43
Changed personal or family plans because of work	2.07 (0.91)	1.80 (0.83)	5.86 (260)	<.001	0.19 to 0.38
Felt frustrated by technology	2.32 (0.95)	1.97 (0.90)	6.18 (258)	<.001	0.24 to 0.46

Aim 2: Engagement With the Intervention and Reactions to Getting Emotional Exhaustion Scores

Of 1575 participants who wrote a baseline gratitude letter, 1007 (63.94%) requested an emailed copy of their gratitude letter and 121 (7.68%) reported having completed this gratitude letter exercise before. At the 1-week follow-up, of 140 participants, 136 (56.7%) answered *yes* to the question "Do you have other people in mind for whom you might write a letter like this?" compared with 46 (19.2%) and 58 (24.2%) who answered *no* or *not sure*, respectively. In addition, 75.4% (181/140) participants agreed slightly, strongly, or very strongly with the statement "Since my first gratitude letter a week ago, it has gotten easier to think of things for which I am grateful," and 45.1% (125/277) participants responded *yes* to the question "Did you talk to anyone about your first gratitude letter?". Independent *t* tests revealed no significant differences in well-being change scores for those who shared their gratitude letters relative to those who did not (EE: t_{266}=1.36, *P*=.18; happiness: t_{270}=−1.04, *P*=.30; and work-life balance: t_{263}=1.04, *P*=.30).

Participants' responses to receiving their EE and work-life balance scores can be found in Table 5. It should be noted that 81.23% (1229/1513) of the respondents agreed slightly or strongly with "knowing my burnout score makes me want to work on it more," and 93.81% (1426/1520) of participants reported wanting to be more resilient.

Those with higher levels of EE at baseline were more likely to agree with all three of these questions (see Table 5). Notably, EE improvers were significantly more likely to agree with the question "Knowing my burnout score makes me want to work on it more" than decliners. Improvers were not more likely than decliners to report being surprised by their scores or a desire to be more resilient.

Table 5. Reactions to receiving burnout feedback, differences by baseline emotional exhaustion levels, and emotional exhaustion improvers vs decliners. Responses were on a 1 (disagree strongly) to 5 (agree strongly) scale.

Burnout feedback item	Severe baseline EE[a] (n=64): those who agreed slightly or strongly, n (%)	Moderate baseline EE (n=515): those who agreed slightly or strongly, n (%)	Mild baseline EE (n=472): those who agreed slightly or strongly, n (%)	No baseline EE (n=473): those who agreed slightly or strongly, n (%)	Concerning EE (n=579) vs no EE (n=473)			EE improvers (n=173) vs decliners (n=62)		
					Mean (SD)	*t (df)*	*P* value	Mean (SD)	*t (df)*	*P* value
"I was surprised by my burnout score"	18 (28.1)	267 (51.8)	211 (44.7)	142 (30)	M1[b] 3.12 (1.38); M2[c] 2.72 (1.26)	4.81 (1036.08)	<.001	M1 3.05 (1.25); M2 3.05 (1.26)	−.01 (233)	.99
"Knowing my burnout score makes me want to work on it more"	51 (79.7)	453 (88.5)	400 (85.5)	322 (67.8)	M1 4.50 (0.84); M2 3.92 (1.11)	9.28 (845.30)	<.001	M1 4.39 (0.91); M2 4.03 (1.16)	2.19 (89.10)	.03
"I would like to be more resilient"	62 (96.8)	490 (95.8)	197 (96)	416 (89.1)	M1 4.82 (0.56); M2 4.50 (0.84)	7.07 (761.00)	<.001	M1 4.74 (0.69); M2 4.56 (0.76)	1.64 (231)	.12

[a]EE: emotional exhaustion.

[b]M1: mean numerical response on a 5-point scale for the concerning EE group (moderate and severe EE) at baseline group.

[c]M2: mean numerical response on a 5-point scale for the no EE at baseline group.

Aim 3: Linguistic Differences Based on Emotional Exhaustion Level at Baseline and Improvements in Emotional Exhaustion

A small number of participants (4.9%) submitted only a few words in the gratitude letter text box. In some cases, it appears that the writer started but did not finish the letter, and in others, the writers engaged in the exercise briefly and superficially (eg, "Mom for raising me. She was amazing."). As word count scores are computed as a percentage of the total number of words, we were concerned that brief letters could bias the linguistic analyses. To reduce potential bias, we excluded letters in the linguistic analyses that were less than 15 words long (77/1575, 4.89% of the sample).

Linguistic category frequencies were compared for those with concerning levels of EE with those without EE at baseline (see Table 6). Negative emotion words were more frequently used in the gratitude letters by those with concerning levels of EE at baseline (P=.005). All other word categories were not used differentially by baseline EE level (P>.05).

Table 6. Independent *t* tests of linguistic differences in gratitude letters by baseline emotional exhaustion (concerning or not) and emotional exhaustion improved (or declined) at the 1-week follow-up.

Hypothesized categories	Baseline EE[a]				EE improvement 1 week later			
	EE concerning, mean (SD)	EE not concerning, mean (SD)	*t (df)*	*P* value	EE decliners (n=63), mean (SD)	EE improvers (n=175), mean (SD)	*t (df)*	*P* value
First person (eg, *I*, *I've*, and *my*)	10.96 (4.52)	10.91 (5.10)	0.17 (1070)	.87	11.88 (4.59)	10.65 (4.24)	−1.93 (236)	.06
First-person plural (eg, *we*, *let's*, and *us*)	0.74 (1.48)	0.84 (1.69)	−1.01 (981.69)	.31	0.82 (1.46)	0.81 (1.65)	−0.03 (236)	.98
Negative emotion (eg, *annoy*, *angry*, and *scream*)	1.16 (1.54)	0.92 (1.36)	2.76 (1066.53)	.005	1.04 (1.56)	1.02 (1.42)	−0.10 (236)	.92
Positive emotion (eg, *appreciate*, *funny*, and *thank*)	8.61 (4.71)	8.61 (5.20)	−0.032 (1070)	.98	9.88 (7.27)	8.51 (4.56)	−1.73 (236)	.09
Cognitive processing (eg, *accept*, *because*, and *realization*)	12.13 (5.28)	11.89 (5.42)	0.73 (1070)	.47	11.52 (5.28)	12.06 (5.27)	0.70 (236)	.48

[a]EE: emotional exhaustion.

The frequencies of word categories were compared for EE improvers vs decliners. There was a statistical trend, albeit nonsignificant, showing that improvers used fewer first-person and positive emotion words (*P*=.06 and .09, respectively). Otherwise, EE improvers vs decliners did not demonstrate different use across the remaining word categories.

Discussion

Principal Findings

This study examined the efficacy of a single-use gratitude letter–writing intervention with a diverse sample of health care workers. Between baseline and the 1-week follow-up, participants reported significant improvements in all three of the well-being measures: EE, happiness, and work-life balance. Participants across both randomized instruction conditions (self-focus vs other focus) reported equivalent improvements in well-being. To our knowledge, this is the first application of a onetime gratitude intervention in health care workers evidencing subsequent improvement in the EE component of burnout.

Participants reported high engagement with the intervention among those who completed the follow-up. Three-fourths of the participants (75.4%) reported that it was easier to think of things to be grateful for since writing their letter. A majority (63.9%) of participants requested an emailed copy of their letter, presumably to reflect on it and/or share with the recipient. Participants equally benefited from the tool regardless of whether they spoke with someone about the letter or not. Although nearly half (45.1%) of the participants reported talking to someone about their letter, we do not know if that person was the letter recipient or not. Previous research has indicated that writing a letter should be followed by a visit to the recipient of the letter [16,17]. It may be that speaking with the letter recipient confers additional benefits, whereas speaking with someone in general (recipient or not) about their letter does not.

Our intervention also gave participants immediate feedback about their EE score during the baseline assessment. Participants were split between being surprised and not being surprised by their scores; however, participants who had higher EE scores were more likely to be surprised. This suggests that a significant portion of health care workers may be burned out, yet they are either not aware of it or are in denial about it [36,37]. After receiving their scores, participants reported very high levels of motivation to reduce their EE (81.2%) and increase their resilience (93.8%). Receiving one's EE score might trigger behavioral change to improve well-being. Indeed, higher agreement to the question "Knowing my burnout score makes me want to work on it more" predicted greater improvement in EE. We cannot determine whether this improvement was because of behavioral change that occurred between the assessments or whether individuals who had higher agreement to this question experienced greater benefit from the gratitude letter–writing exercise or because of some other reason. However, considerable research has shown that greater awareness of a personal problem is a key step in behavioral change [38,39]. A recent meta-analysis of burnout prevalence in 109,628 physicians across 182 studies suggests that many physicians are answering burnout questions, but we believe that the proportion receiving feedback about their burnout is close to 0 [40]. It could be that burnout psychoeducation on its own may not be sufficiently motivating. Rather, receiving one's burnout score can be likened to someone *stepping on the scale* before engaging in more physical activity, bringing a personal awareness of his or her status to trigger behavioral change.

Linguistic analyses revealed that participants with moderate and severe EE at baseline used more negative emotion words in their letters compared with participants without EE. Similarly, previous research has shown that depressives use more negative emotion words in their writing samples [19]. Surprisingly, we did not find differences in the use of first-person singular or plural, positive emotions, or cognitive processing word categories by EE severity. Although we found a statistical trend associating greater reductions in EE with the use of fewer first-person singular words (eg, *I* and me) and fewer positive emotion words, neither of them reached statistical significance (*P*=.06 and *P*=.09, respectively; see the Multimedia Appendix

1). Differences in these categories have been found in depressives vs controls in prior research, albeit with very different writing prompts [19]. Writing a note of gratitude might draw out a different pattern of language compared with the expressive writing prompts often used by other linguistics researchers, which ask participants to write about their deepest emotions and emotional challenges over a 4-day period [41]. Thus, linguistic patterns in the gratitude letter tool may not reflect psychological well-being as clearly as was found in other studies. Given that the gratitude tool appears to be beneficial, it may be that an overall feeling of gratitude induced by this intervention drives improvement, rather than a complex linguistic pattern across word categories. In other words, it may simply be more important *that* you are grateful rather than *how* you are grateful.

In line with prior research, we expected that participants who were assigned to write more about themselves, rather than the recipient of the letter, would benefit more from the exercise. Although our prompts appeared to change the linguistic focus (ie, the self-focused condition used more *I* and less *me* language than the other-focused condition), one condition did not further enhance well-being relative to the other. Notably, prior research on the benefits of self-focused language examined verbal expressions of gratitude, rather than written letters [18]. It may be that in a letter format, the impact of focusing on oneself during expressions of gratitude is inherently weaker or diluted by the additional time spent writing.

This study and prior research indicate that the gratitude letter tool, and similar positive psychology tools, might improve well-being by shifting one's focus from what is *not* going well to what *is* going well. Indeed, the negativity bias makes people significantly more likely to pay attention to negative vs positive stimuli [42]. Individuals with higher levels of negativity bias are more likely to be depressed [43]. When prompted to positively reflect, participants appear to consider good things they had previously taken for granted, had not thought much about, or were formerly overshadowed by attention to negative experiences. By intentionally shifting attention to what is going well, participants may recalibrate their perceptions toward noticing more positive experiences and be primed to engage in other self-care activities such as exercising [44]. Indeed, 75.4% of follow-up participants reported that it became easier to think of things to be grateful for after writing their first gratitude letter.

We chose to implement gratitude letter writing as a web-based intervention to increase the availability of the tool and increase our sample size. In fact, our baseline sample size was larger than we anticipated. However, it is possible that a web-based format made it easier for participants to ignore the 1-week follow-up survey message. Web-based studies typically have high dropout rates, likely because of the ease of ignoring the request and reduction of interpersonal reciprocity norms. It is possible that personal characteristics make web-based tools more appealing to some, and this could have influenced enrollment and engagement in the study. Future research should explore what characteristics predispose individuals to benefit from web-based vs in-person well-being tools.

Limitations

Our study has several limitations. First, our study included randomization to two letter conditions, but there was no nonintervention control group. In addition, the intervention included both the gratitude letter activity and receiving one's baseline EE score, component effects of which we cannot disentangle. These aspects preclude our establishing a causal relationship between gratitude letter writing and improvement in well-being. However, prior randomized controlled gratitude interventions (eg, nightly gratitude log and gratitude visit) have demonstrated improvement compared with placebo controls, which bolsters confidence in the results of this study [16].

Nonetheless, the observed benefits of this intervention may in part arise from expectancy bias or placebo effects. We observe that participants who engaged in the gratitude letter tool and completed the 1-week follow-up reported significant improvements across all outcomes. It is possible that there were intrinsic characteristics of the cohort that completed follow-up assessments that predisposed them to improvement, whereas participants not completing postintervention assessments may not have benefited to the same degree.

Selection bias comprises another noted limitation. Our sample largely comprised women (69.9% of participants at baseline). Nevertheless, this approximates the base rates of women working in health care in the United States (in 2017, 75% of health care workers were female) [45]. Future research should employ a placebo-controlled design, examine the gratitude letter or EE feedback alone, and potentially oversample for men, all of which will serve to better determine causality, divine specific effects of the tool, and establish the extent of efficacy across genders. Overall, the high level of engagement observed within our cohort, significant participant interest in discussing letters, and self-reported improvements in the ability to identify that for which they were grateful suggest a meaningful impact of this short letter-writing activity.

Our study also suffered from what looks like attrition, with 18% of participants who completed a baseline assessment also completing the 1-week follow-up. A significant component likely arose from varied means by which health care workers were referred to the tool. In some cases, participants received a brief link at the conclusion of a talk or meeting, without mention of any follow-up evaluation: "Consider giving this just 7 minutes to see what you think of a bite-sized resilience activity." In other words, although many participants enrolled knowing that there would be a 1-week follow-up, some participants likely never intended on completing the follow-up at the outset. It is possible that individuals who failed to complete the follow-up did not benefit from the tool to the same extent as those who did complete it, overinflating the tool's effects. That said, our attrition rates are similar to comparable studies [46-48]. Attrition is a primary barrier to evaluating web-based interventions, with levels often reaching 60% to 80%. Among people seeking treatment for obesity using weight loss programs in medical centers, one-third to half discontinue their program and are lost to follow-up [49,50]. Similarly, over 40% of people seeking treatment through a smoking cessation clinic were lost to follow-up [51], and 48% of web-based

smoking cessation individuals were lost to follow-up [52]. Even when offered significant financial incentives (US $600) to participate in a cessation program, only 20% of those invited chose to participate [53]. In future iterations of this tool, we plan to improve the follow-up rates by setting clear expectations for receiving and completing follow-up assessments, identifying whether participants correctly entered their contact information into our system, and implementing reminders.

Perhaps most importantly, the gratitude letter tool is designed to build participants' personal well-being resources but does not address the demands placed upon them in their work environments. Ultimately, the pace and intensity of contemporary health care are major unaddressed sources of strain that impair well-being. The gratitude letter tool, although promising, should be part of a much larger toolkit of valid and accessible interventions facilitating individual well-being and institutional changes.

Despite these limitations, this study finds that the gratitude letter tool offers a simple, brief, and free individual intervention to improve health care workers' EE, happiness, and work-life balance. The flexibility of this single-use tool is highly desirable amid the fast pace of health care. Indeed, it can be inserted into a staff meeting, during a break, or completed briefly at home. In an era of alarmingly high levels of health care worker EE, the gratitude letter tool appears to be a promising arrow in the quiver to improve well-being.

Conclusions

A single-exposure gratitude letter–writing intervention is a promising tool to improve EE, happiness, and work-life balance in health care workers. There is preliminary evidence that receiving feedback about your EE serves to motivate further well-being action and, in turn, improve EE. The gratitude letter tool [23] is brief, simple, and free and requires no further action after writing a 7-min letter, which are all important factors for busy health care workers and overburdened health care systems. Given the relationship between health care workers' well-being and health care quality, engaging in the gratitude letter tool may help relieve EE and, in turn, improve patient safety and quality.

Acknowledgments

This work was supported by the National Institutes of Health grant number R01 HD084679-01. The authors would like to express their deep gratitude to Christen Noratel for her tireless efforts over the past 10 years in supporting trainings and seminars that shared the gratitude letter tool with health care workers.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Discussion on trending linguistics results.
[DOCX File , 14 KB - jmir_v22i5e15562_app1.docx]

References

1. Aiken LH, Sermeus W, van den Heede K, Sloane DM, Busse R, McKee M, et al. Patient safety, satisfaction, and quality of hospital care: cross sectional surveys of nurses and patients in 12 countries in Europe and the United States. Br Med J 2012 Mar 20;344:e1717 [FREE Full text] [doi: 10.1136/bmj.e1717] [Medline: 22434089]
2. Rothenberger DA. Physician burnout and well-being: a systematic review and framework for action. Dis Colon Rectum 2017 Jun;60(6):567-576. [doi: 10.1097/DCR.0000000000000844] [Medline: 28481850]
3. Shanafelt TD, Boone S, Tan L, Dyrbye LN, Sotile W, Satele D, et al. Burnout and satisfaction with work-life balance among US physicians relative to the general US population. Arch Intern Med 2012 Oct 8;172(18):1377-1385. [doi: 10.1001/archinternmed.2012.3199] [Medline: 22911330]
4. McCray LW, Cronholm PF, Bogner HR, Gallo JJ, Neill RA. Resident physician burnout: is there hope? Fam Med 2008 Oct;40(9):626-632 [FREE Full text] [Medline: 18830837]
5. Wheeler DL, Vassar M, Worley JA, Barnes LL. A reliability generalization meta-analysis of coefficient alpha for the Maslach Burnout Inventory. Educ Psychol Meas 2011;71(1):231-244. [doi: 10.1177/0013164410391579]
6. West CP, Tan AD, Habermann TM, Sloan JA, Shanafelt TD. Association of resident fatigue and distress with perceived medical errors. J Am Med Assoc 2009 Sep 23;302(12):1294-1300. [doi: 10.1001/jama.2009.1389] [Medline: 19773564]
7. Dyrbye LN, Burke SE, Hardeman RR, Herrin J, Wittlin NM, Yeazel M, et al. Association of clinical specialty with symptoms of burnout and career choice regret among US resident physicians. JAMA 2018 Sep 18;320(11):1114-1130 [FREE Full text] [doi: 10.1001/jama.2018.12615] [Medline: 30422299]
8. Dyrbye LN, Thomas MR, Massie FS, Power DV, Eacker A, Harper W, et al. Burnout and suicidal ideation among US medical students. Ann Intern Med 2008 Sep 2;149(5):334-341. [doi: 10.7326/0003-4819-149-5-200809020-00008] [Medline: 18765703]

9. Krasner MS, Epstein RM, Beckman H, Suchman AL, Chapman B, Mooney CJ, et al. Association of an educational program in mindful communication with burnout, empathy, and attitudes among primary care physicians. J Am Med Assoc 2009 Sep 23;302(12):1284-1293. [doi: 10.1001/jama.2009.1384] [Medline: 19773563]
10. Sexton JB, Schwartz SP, Chadwick WA, Rehder KJ, Bae J, Bokovoy J, et al. The associations between work-life balance behaviours, teamwork climate and safety climate: cross-sectional survey introducing the work-life climate scale, psychometric properties, benchmarking data and future directions. BMJ Qual Saf 2017 Aug;26(8):632-640 [FREE Full text] [doi: 10.1136/bmjqs-2016-006032] [Medline: 28008006]
11. Schwartz SP, Adair KC, Bae J, Rehder KJ, Shanafelt TD, Profit J, et al. Work-life balance behaviours cluster in work settings and relate to burnout and safety culture: a cross-sectional survey analysis. BMJ Qual Saf 2019 Feb;28(2):142-150 [FREE Full text] [doi: 10.1136/bmjqs-2018-007933] [Medline: 30309912]
12. Emmons RA, Stern R. Gratitude as a psychotherapeutic intervention. J Clin Psychol 2013 Aug;69(8):846-855. [doi: 10.1002/jclp.22020] [Medline: 23775470]
13. Toepfer SM, Cichy K, Peters P. Letters of gratitude: further evidence for author benefits. J Happiness Stud 2012;13(1):187-201. [doi: 10.1007/s10902-011-9257-7]
14. O'Connell BH, O'Shea D, Gallagher S. Examining psychosocial pathways underlying gratitude interventions: a randomized controlled trial. J Happiness Stud 2018;19(8):2421-2444. [doi: 10.1007/s10902-017-9931-5]
15. Southwell S, Gould E. A randomised wait list-controlled pre–post–follow-up trial of a gratitude diary with a distressed sample. J Posit Psychol 2017;12(6):579-593. [doi: 10.1080/17439760.2016.1221127]
16. Emmons RA, McCullough ME. Counting blessings versus burdens: an experimental investigation of gratitude and subjective well-being in daily life. J Pers Soc Psychol 2003 Feb;84(2):377-389. [doi: 10.1037//0022-3514.84.2.377] [Medline: 12585811]
17. Seligman ME, Steen TA, Park N, Peterson C. Positive psychology progress: empirical validation of interventions. Am Psychol 2005;60(5):410-421. [doi: 10.1037/0003-066X.60.5.410] [Medline: 16045394]
18. Algoe SB, Kurtz LE, Hilaire NM. Putting the 'You' in 'Thank You': examining other-praising behavior as the active relational ingredient in expressed gratitude. Soc Psychol Personal Sci 2016 Sep;7(7):658-666 [FREE Full text] [doi: 10.1177/1948550616651681] [Medline: 27570582]
19. Rude SS, Gortner EM, Pennebaker JW. Language use of depressed and depression-vulnerable college students. Cogn Emot 2004 Dec;18(8):1121-1133. [doi: 10.1080/02699930441000030]
20. Eichstaedt JC, Smith RJ, Merchant RM, Ungar LH, Crutchley P, Preoţiuc-Pietro D, et al. Facebook language predicts depression in medical records. Proc Natl Acad Sci U S A 2018 Oct 30;115(44):11203-11208 [FREE Full text] [doi: 10.1073/pnas.1802331115] [Medline: 30322910]
21. Kleim B, Horn AB, Kraehenmann R, Mehl MR, Ehlers A. Early linguistic markers of trauma-specific processing predict post-trauma adjustment. Front Psychiatry 2018;9:645 [FREE Full text] [doi: 10.3389/fpsyt.2018.00645] [Medline: 30568607]
22. Adair KC, Quow K, Frankel A, Mosca PJ, Profit J, Hadley A, et al. The Improvement Readiness scale of the SCORE survey: a metric to assess capacity for quality improvement in healthcare. BMC Health Serv Res 2018 Dec 17;18(1):975 [FREE Full text] [doi: 10.1186/s12913-018-3743-0] [Medline: 30558593]
23. Duke Qualtrics. Gratitude Letter URL: https://duke.qualtrics.com/jfe/form/SV_200ktTwZpgGTi6x?Q_JFE=qdg [accessed 2020-01-23]
24. Sexton JB, Adair KC. Forty-five good things: a prospective pilot study of the Three Good Things well-being intervention in the USA for healthcare worker emotional exhaustion, depression, work-life balance and happiness. BMJ Open 2019 Mar 20;9(3):e022695 [FREE Full text] [doi: 10.1136/bmjopen-2018-022695] [Medline: 30898795]
25. Maslach C, Jackson SE, Leiter MP. Maslach Burnout Inventory Manual. Third Edition. Palo Alto, CA: Consulting Psychologists Press; 1996.
26. Loera B, Converso D, Viotti S. Evaluating the psychometric properties of the Maslach Burnout Inventory-Human Services Survey (MBI-HSS) among Italian nurses: how many factors must a researcher consider? PLoS One 2014;9(12):e114987 [FREE Full text] [doi: 10.1371/journal.pone.0114987] [Medline: 25501716]
27. Kleijweg JH, Verbraak MJ, van Dijk MK. The clinical utility of the Maslach Burnout Inventory in a clinical population. Psychol Assess 2013 Jun;25(2):435-441. [doi: 10.1037/a0031334] [Medline: 23356679]
28. Schaufeli WB, Bakker AB, Hoogduin K, Schaap C, Kladler A. On the clinical validity of the Maslach Burnout Inventory and the burnout measure. Psychol Health 2001 Sep;16(5):565-582. [doi: 10.1080/08870440108405527] [Medline: 22804499]
29. Sexton JB, Adair KC, Leonard MW, Frankel TC, Proulx J, Watson SR, et al. Providing feedback following Leadership WalkRounds is associated with better patient safety culture, higher employee engagement and lower burnout. BMJ Qual Saf 2018 Apr;27(4):261-270 [FREE Full text] [doi: 10.1136/bmjqs-2016-006399] [Medline: 28993441]
30. Lyubomirsky S, Lepper HS. A measure of subjective happiness: preliminary reliability and construct validation. Soc Indic Res 1999;46(2):137-155. [doi: 10.1023/A:1006824100041]
31. Lin JD, Lin PY, Wu CL. Wellbeing perception of institutional caregivers working for people with disabilities: use of Subjective Happiness Scale and Satisfaction with Life Scale analyses. Res Dev Disabil 2010;31(5):1083-1090. [doi: 10.1016/j.ridd.2010.03.009] [Medline: 20400262]

32. Adair K, Sexton J. 3 Good Tools: Positively reflecting backwards and forwards is associated with robust improvements in well-being across 3 distinct interventions. Conditionally accepted at the Journal of Positive Psychology 2019:9999999-9999999.
33. Peterson RA. A meta-analysis of Cronbach's coefficient alpha. J Consum Res 1994 Sep;21(2):381-391. [doi: 10.1086/209405]
34. Pennebaker JW, Boyd RL, Jordan K, Blackburn K. Repository Home. Austin, TX: University of Texas at Austin; 2015. The Development and Psychometric Properties of LIWC2015 URL: https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf [accessed 2020-01-21]
35. Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. J Lang Soc Psychol 2010;29(1):24-54. [doi: 10.1177/0261927X09351676]
36. Gazelle G, Liebschutz JM, Riess H. Physician burnout: coaching a way out. J Gen Intern Med 2015 Apr;30(4):508-513 [FREE Full text] [doi: 10.1007/s11606-014-3144-y] [Medline: 25527340]
37. Doolittle BR, Windish DM, Seelig CB. Burnout, coping, and spirituality among internal medicine resident physicians. J Grad Med Educ 2013 Jun;5(2):257-261 [FREE Full text] [doi: 10.4300/JGME-D-12-00136.1] [Medline: 24404269]
38. Prochaska JO, Prochaska JM. Why don't continents move? Why don't people change? J Psychother Integr 1999;9(1):83-102 [FREE Full text] [doi: 10.1023/A:1023210911909]
39. Norcross JC, Krebs PM, Prochaska JO. Stages of change. J Clin Psychol 2011 Feb;67(2):143-154. [doi: 10.1002/jclp.20758] [Medline: 21157930]
40. Rotenstein LS, Torre M, Ramos MA, Rosales RC, Guille C, Sen S, et al. Prevalence of burnout among physicians: a systematic review. J Am Med Assoc 2018 Sep 18;320(11):1131-1150 [FREE Full text] [doi: 10.1001/jama.2018.12777] [Medline: 30326495]
41. Pennebaker JW, Chung CK. Expressive writing: connections to physical and mental health. In: Friedman HS, editor. The Oxford Handbook of Health Psychology. New York, NY: Oxford University Press; 2011:417-437.
42. Baumeister RF, Bratslavsky E, Finkenauer C, Vohs KD. Bad is stronger than good. Rev Gen Psychol 2001;5(4):323-370. [doi: 10.1037/1089-2680.5.4.323]
43. Gollan JK, Hoxha D, Hunnicutt-Ferguson K, Norris CJ, Rosebrock L, Sankin L, et al. Twice the negativity bias and half the positivity offset: evaluative responses to emotional information in depression. J Behav Ther Exp Psychiatry 2016 Sep;52:166-170 [FREE Full text] [doi: 10.1016/j.jbtep.2015.09.005] [Medline: 26434794]
44. Lyubomirsky S, Layous K. How do simple positive activities increase well-being? Curr Dir Psychol Sci 2013;22(1):57-62. [doi: 10.1177/0963721412469809]
45. United States Department of Labor. US Bureau of Labor Statistics Latest Numbers. Employed persons by detailed occupation, sex, race, and Hispanic or Latino ethnicity URL: https://www.bls.gov/cps/cpsaat11.htm [accessed 2018-08-03]
46. Abbott JA, Klein B, Hamilton C, Rosenthal AJ. The impact of online resilience training for sales managers on wellbeing and performance. Sensoria 2009;5(1):89-95. [doi: 10.7790/ejap.v5i1.145]
47. Gander F, Proyer RT, Ruch W, Wyss T. Strength-based positive interventions: further evidence for their potential in enhancing well-being and alleviating depression. J Happiness Stud 2012;14(4):1241-1259 [FREE Full text] [doi: 10.1007/s10902-012-9380-0]
48. Mongrain M, Anselmo-Matthews T. Do positive psychology exercises work? A replication of Seligman et al. (2005). J Clin Psychol 2012 Apr;68(4):382-389. [doi: 10.1002/jclp.21839] [Medline: 24469930]
49. Norcross JC, Mrykalo MS, Blagys MD. Auld lang syne: success predictors, change processes, and self-reported outcomes of New Year's resolvers and nonresolvers. J Clin Psychol 2002 Apr;58(4):397-405. [doi: 10.1002/jclp.1151] [Medline: 11920693]
50. Wing RR, Phelan S. Long-term weight loss maintenance. Am J Clin Nutr 2005 Jul;82(1 Suppl):222S-225S. [doi: 10.1093/ajcn/82.1.222S] [Medline: 16002825]
51. Bahadir A, Iliaz S, Yurt S, Ortakoylu MG, Bakan ND, Yazar E. Factors affecting dropout in the smoking cessation outpatient clinic. Chron Respir Dis 2016 May;13(2):155-161 [FREE Full text] [doi: 10.1177/1479972316629953] [Medline: 26846679]
52. Geraghty AW, Torres LD, Leykin Y, Pérez-Stable EJ, Muñoz RF. Understanding attrition from international Internet health interventions: a step towards global eHealth. Health Promot Int 2013 Sep;28(3):442-452. [doi: 10.1093/heapro/das029] [Medline: 22786673]
53. Halpern SD, Harhay MO, Saulsgiver K, Brophy C, Troxel AB, Volpp KG. A pragmatic trial of e-cigarettes, incentives, and drugs for smoking cessation. N Engl J Med 2018 Jun 14;378(24):2302-2310. [doi: 10.1056/NEJMsa1715757] [Medline: 29791259]

Abbreviations

EE: emotional exhaustion
LIWC: Linguistic Inquiry and Word Count
MBI: Maslach Burnout Inventory
SHS: Subjective Happiness Scale

Edited by G Eysenbach; submitted 19.07.19; peer-reviewed by S Tao, J Brož, M Lozano-Lozano, Y Bartlett; comments to author 19.10.19; revised version received 31.10.19; accepted 16.12.19; published 14.05.20.

Please cite as:
Adair KC, Rodriguez-Homs LG, Masoud S, Mosca PJ, Sexton JB
Gratitude at Work: Prospective Cohort Study of a Web-Based, Single-Exposure Well-Being Intervention for Health Care Workers
J Med Internet Res 2020;22(5):e15562
URL: https://www.jmir.org/2020/5/e15562
doi:10.2196/15562
PMID:32406864

Original Paper

Web-Based Self-Management for Patients With Lymphoma: Assessment of the Reach of Intervention of a Randomized Controlled Trial

Lindy P J Arts[1], MSc; Simone Oerlemans[1,2], PhD; Eduardus F M Posthuma[3], MD, PhD; Djamila E Issa[4], MD, PhD; Margriet Oosterveld[5], MD, PhD; René van der Griend[6], MD, PhD; Marten R Nijziel[7], MD, PhD; Lonneke V van de Poll-Franse[1,2,8], PhD

[1]Netherlands Comprehensive Cancer Organisation, Utrecht, Netherlands

[2]Center of Research on Psychology in Somatic Diseases, Department of Medical and Clinical Psychology, Tilburg University, Tilburg, Netherlands

[3]Department of Internal Medicine, Reinier de Graaf Hospital, Delft, Netherlands

[4]Department of Internal Medicine, Jeroen Bosch Hospital, 's-Hertogenbosch, Netherlands

[5]Department of Internal Medicine, Canisius-Wilhelmina Hospital, Nijmegen, Netherlands

[6]Department of Internal Medicine, Diakonessenhuis, Utrecht/Zeist, Netherlands

[7]Department of Internal Medicine, Catharina Hospital, Eindhoven, Netherlands

[8]Division of Psychosocial Research and Epidemiology, Netherlands Cancer Institute, Amsterdam, Netherlands

Corresponding Author:
Simone Oerlemans, PhD
Netherlands Comprehensive Cancer Organisation
Godebaldkwartier 419
Utrecht, 3511DT
Netherlands
Phone: 31 882346000
Email: s.oerlemans@iknl.nl

Abstract

Background: Randomized controlled trials (RCTs) often provide accurate estimates of the internal validity of an intervention but lack information on external validity (generalizability). We conducted an RCT on the effectiveness of a self-management intervention among patients with lymphoma in a population-based setting.

Objective: The objectives of the current study were to describe the proportion of RCT participants compared to all patients invited to participate, and compare sociodemographic and clinical characteristics of RCT participants with all respondents, all patients invited to participate, and all patients selected from the Netherlands Cancer Registry (NCR) to determine the reach of the intervention. An additional objective was to assess differences on RCT outcome variables between RCT and paper respondents.

Methods: Patients with lymphoma or chronic lymphocytic leukemia ≥18 years old at diagnosis from 13 hospitals in the Netherlands were selected from the population-based NCR, which routinely collects data on sociodemographic and clinical characteristics. Eligible patients were invited to participate in an RCT and complete a questionnaire. Web-based completion determined RCT enrollment, whereas paper respondents were followed observationally.

Results: A total of 1193 patients were selected from the NCR, 892 (74.77%) of whom were invited to participate in the trial by their hematologist after verifying eligibility. Among those invited, 25.4% (227/892) completed the web-based questionnaire and were enrolled in the RCT. The RCT participants were younger and there was a higher proportion of men than nonparticipants (P<.001). In addition, 25.7% (229/892) of those invited opted to participate in the paper-based observational follow-up study. Compared with paper respondents, RCT participants were younger (P<.001), with a higher proportion of men (P=.002), and had higher education levels (P=.02). RCT participants more often wanted to receive all available information on their disease (P<.001), whereas paper respondents reported higher levels of emotional distress (P=.009).

Conclusions: From a population-based sample of eligible patients, the participation rate in the RCT was approximately 25%. RCT participants may not be representative of the target population because of different sociodemographic and clinical characteristics. Since RCT participants represent a minority of the target population, RCT results should be interpreted with caution as patients in the RCT may be those least in need of a self-management intervention.

Trial Registration: Netherlands Trial Register NTR5953; https://www.trialregister.nl/trial/5790

(J Med Internet Res 2020;22(5):e17018) doi:10.2196/17018

KEYWORDS

reach; uptake; participation, web-based intervention; pragmatic; randomized controlled trial; population-based registry; lymphoma

Introduction

Randomized controlled trials (RCTs) are widely considered to be the gold standard for evaluating the effects of an intervention in behavioral and psycho-oncological research [1,2]. In contrast to the effects of interventions that are most often examined extensively in RCTs [3], much less attention has been paid to the proportion of patients who participate in these interventions and whether those who choose to participate are representative of the target population in terms of sociodemographic and clinical characteristics [2,4]. Thus, RCTs often provide accurate estimates of the internal validity (ie, effect of an intervention for the sample enrolled in the RCT), but do not typically provide information about the external validity or generalizability (ie, effect of an intervention in the target population) [5-7].

The reach of an RCT provides information on the absolute number, proportion, and representativeness of the sample that participates in the trial [8]. The absolute number and proportion of RCT participants are relatively easy to assess and are therefore most often reported. However, few studies report the representativeness of the sample enrolled in an RCT, which is a much more challenging metric to assess [8,9] since it requires sociodemographic information, and preferable psychosocial, clinical, or case mix information, on RCT participants as well as nonparticipants. It is particularly challenging to collect information on nonparticipants who typically do not consent to be included in the research [8].

Interventions with promising effects in RCTs have been implemented in daily practice without specific knowledge of the generalizability of the results. Therefore, more attention should be paid to providing information related to the representativeness of the sample enrolled in an RCT. Lack of representativeness may occur as a result of inadequate selection procedures (ie, sampling bias) or when the probability of nonparticipation in the study is related to the object of research (ie, nonresponse bias) [10,11].

To fill this gap, the aim of the current study was to address the reach of a web-based self-management intervention within the context of the Lymphoma InterVEntion (LIVE) trial (Netherlands Trial Register NTR5953), whose objectives have been described elsewhere [12]. For the LIVE trial, patients were selected from the population-based Netherlands Cancer Registry (NCR) that routinely collects data on sociodemographic and clinical characteristics. This can provide unique insight into the characteristic differences between RCT participants and nonparticipants to estimate the reach of this intervention. The primary objectives were to (1) describe the proportion of RCT participants compared to all patients invited to participate, and (2) compare sociodemographic and clinical characteristics of RCT participants with those of all respondents (ie, patients who completed a web-based or paper questionnaire), all patients invited to participate, and all patients selected from the NCR. In addition, as patients had the option of completing a web-based questionnaire (ie, enrollment in the RCT) or a paper-based questionnaire (ie, no enrollment in the RCT; observational cohort), a secondary objective was to assess baseline differences in psychological distress, self-management skills, and satisfaction with information provision (ie, RCT outcome variables) between the two groups.

Methods

Study Design

Baseline data were collected from an RCT embedded in a population-based registry [13] as an observational cross-sectional dataset without information on the effectiveness of the intervention. In short, the LIVE trial examines the effectiveness of feedback on patient-reported outcomes and a web-based self-management intervention on self-management skills, satisfaction with information provision, and psychological distress among patients with lymphoma [12].

Participants and Recruitment Procedure

From October 2016 to February 2019, patients who were diagnosed with lymphoma, including Hodgkin lymphoma, nonHodgkin lymphoma, or chronic lymphocytic leukemia (CLL) as defined by the International Classification of Diseases for Oncology-3 codes [14], from 13 hospitals in the Netherlands were selected for participation via the NCR. The NCR registers all patients newly diagnosed with cancer in the Netherlands within the first year after diagnosis and routinely collects detailed data on sociodemographic and clinical characteristics (eg, patient age and sex, date of cancer diagnosis, cancer type, and primary treatment). Patients had to be 18 years or older at the time of study invitation. Treating hematologists were asked to verify the patients' eligibility for the study and to exclude patients who were deceased, had severe psychopathology, were too ill, were not able to complete a questionnaire in Dutch, or had severe cognitive impairment. All eligible patients were invited by mail to participate by their own hematologist. Patients had the option to complete a web-based or paper-based questionnaire. Patients were informed that completion of the web-based questionnaire automatically resulted in enrollment in the RCT with randomization to one of the study arms, whereas completion of a paper questionnaire resulted in participation in the observational Patient Reported Outcomes Following Initial treatment and Long term Evaluation of Survivorship (PROFILES) registry [13] but not enrollment in the RCT. To address the primary objective of the current study—describing the proportion of RCT participants compared to all patients invited to participate—paper respondents were

assessed as nonparticipants as they did not participate in the RCT.

Measures

Sociodemographic and Clinical Measures

Sociodemographic characteristics (age and sex) and detailed clinical information (date of diagnosis, cancer type, primary treatment) were available from the NCR. Information on education level and marital status was assessed from the questionnaire (data only available for respondents).

Comorbidities at the time of questionnaire completion were assessed using an adapted version of the self-administered comorbidity questionnaire [15]. Patients were asked to identify comorbidities present within the past 12 months, including heart disease, hypertension, arthritis, stroke, lung disease, diabetes, stomach disease, kidney disease, liver disease, anemia, thyroid disease, and rheumatoid arthritis. Positive responses were summed to a total score ranging from 0 to 12 (data only available for respondents).

Personality Traits

Personality traits were assessed with the Big Five Inventory, a 44-item inventory designed to measure the Big Five dimensions of personality: extraversion, neuroticism, conscientiousness, agreeableness, and openness to experience [16]. Each item was scored on a 5-point scale. Scale scores were obtained by averaging all items for each domain ranging from 0 to 5. Each trait is assumed to represent a continuum from high to low on the specific attribute and is partnered with a trait on the opposite pole of the spectrum [17,18].

Information Preferences

One question from an adapted version of the Information Satisfaction Questionnaire was used to measure information preferences [19]. Patients had to categorize themselves into one of three groups: those who would like (1) all available information, (2) only positive information about the illness, and (3) only limited information. Patients were further asked whether they use the internet (yes/no).

Psychological Distress, Self-Management Skills, and Satisfaction With Information Provision

Primary outcomes to assess the effectiveness of the intervention were psychological distress, self-management skills, and satisfaction with information provision.

Psychological distress was assessed with the 14-item Hospital Anxiety and Depression Scale [20]. Each item is rated on a 4-point scale from 0 to 3. The total score was obtained by adding all item scores and ranged from 0 to 42, in which higher scores indicate higher levels of psychological distress [21]. Patients with a Hospital Anxiety and Depression Scale sum score ≥13 were categorized as “psychologically distressed” [22].

Self-management skills were assessed with the Health Education Impact Questionnaire (heiQ) that contains 40 items across eight scales: positive and active engagement in life, health-directed activities, skill and technique acquisition, constructive attitudes and approaches, self-monitoring and insight, health service navigation, social integration and support, and emotional distress [23]. Each item is scored on a 4-point scale. Scale scores were obtained by averaging all items for each domain and ranged from 1 to 4. Higher scores indicate better status or self-management, except for emotional distress in which higher scores indicate greater distress [23].

Satisfaction with overall information provision was assessed with one item from an adapted version of the Information Satisfaction Questionnaire [19]. Patients were asked to rate their level of satisfaction for overall information provision on a scale from 1 (“very unsatisfied”) to 5 (“very satisfied”).

Statistical Analyses

The proportion of RCT participants (ie, participation rate) was calculated by dividing the number of patients who were enrolled in the RCT by the total number of eligible patients who were invited to participate. Sociodemographic and clinical characteristics of RCT participants were compared with those of all respondents (ie, patients who completed a web-based or paper questionnaire), all patients invited to participate, and all patients selected from the population-based NCR. In addition, personality traits and information preferences of RCT participants were compared with those of all respondents. Differences on sociodemographic and clinical characteristics between RCT participants and nonparticipants were compared using analysis of variance for continuous variables and Chi-square tests for categorical variables.

Differences in baseline psychological distress, self-management skills, and satisfaction with information provision (ie, RCT outcome variables) between RCT participants and paper respondents were compared using analysis of variance for continuous variables and Chi-square tests for categorical variables. All statistical analyses were performed with SAS version 9.4 (Cary, NC, USA). $P \leq .05$ indicated statistically significant differences.

Results

Patients Selected From the Netherlands Cancer Registry

As shown in Figure 1, a total of 1193 patients with lymphoma or CLL who were ≥18 years old at diagnosis from 13 hospitals were selected from the population-based NCR. The basic characteristics of the invited patients are summarized in Table 1. The majority of patients were men, were diagnosed with high-grade nonHodgkin lymphoma, and were actively being treated, with chemotherapy as the most common treatment.

Figure 1. Flowchart of the data collection process. CLL: chronic lymphocytic leukemia; NCR: Netherlands Cancer Registry; RCT: randomized controlled trial; PROFILES: Patient Reported Outcomes Following Initial treatment and Long term Evaluation of Survivorship.

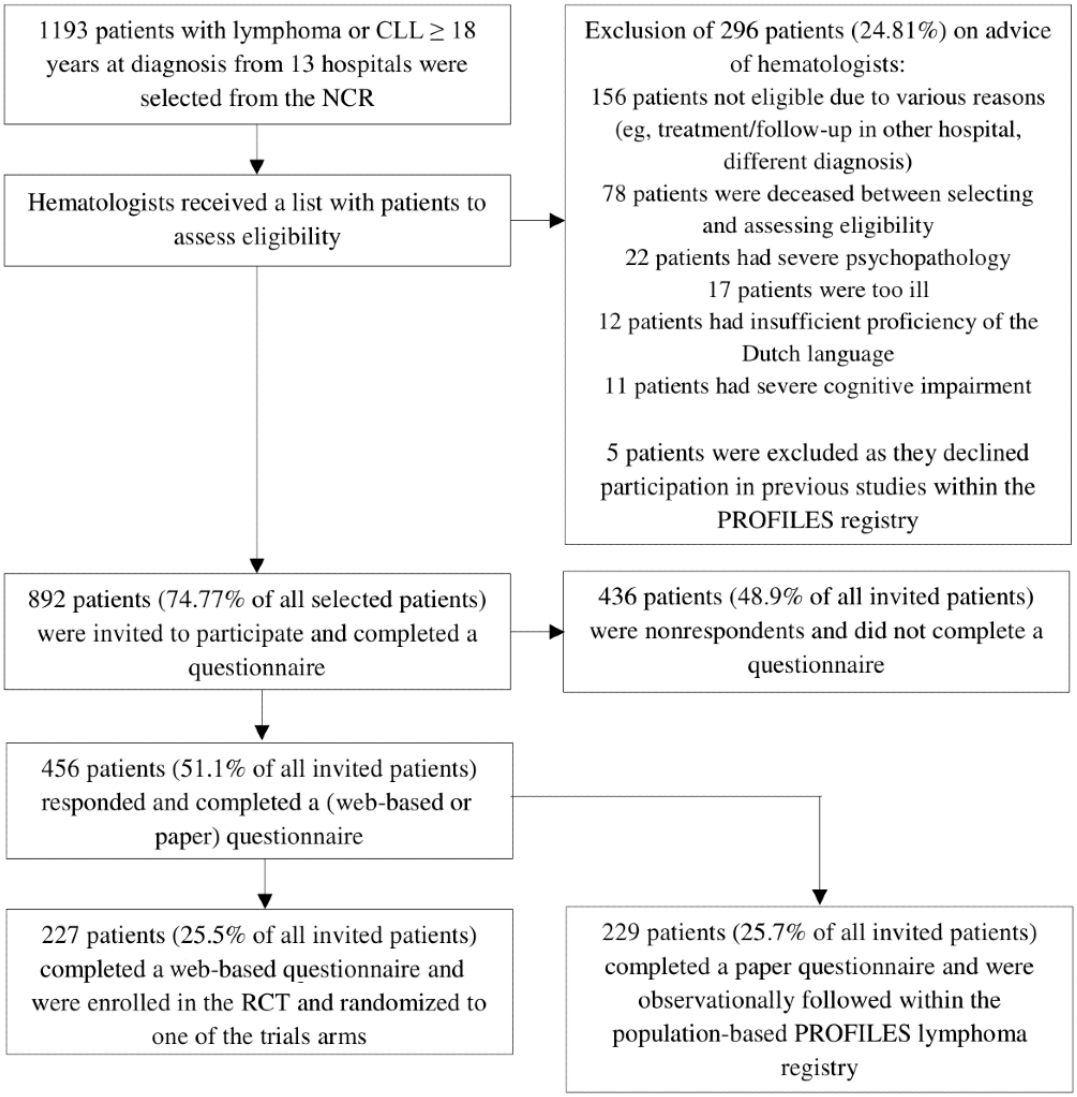


Table 1. Baseline characteristics of patients selected from the Netherlands Cancer Registry (NCR) according to participation[a].

Characteristic	Patients selected from the NCR (N=1193)	Patients invited to participate (n=892)	Respondents (n=456)	RCT[b] participants (n=227)
Sociodemographic characteristics, n (%)				
Age at time of survey (years), mean (SD)	64.7 (15.6)	64.1 (15.3)	64.5 (13.5)	60.7 (13.4)
Sex, n (%)				
Male	725 (60.8)	537 (60.2)	291 (63.8)	161 (70.9)
Female	468 (39.2)	355 (39.8)	165 (36.2)	66 (29.1)
Education, n (%)				
Low[c]	N/A[d]	N/A	33 (7.2)	6 (2.6)
Medium[e]	N/A	N/A	260 (57.0)	106 (46.7)
High[f]	N/A	N/A	159 (34.9)	114 (50.2)
Partner (yes), n (%)	N/A	N/A	355 (77.9)	190 (83.7)
Clinical characteristics				
Months since diagnosis, mean (SD)	13.9 (3.5)	14.0 (3.4)	14.2 (3.3)	14.0 (3.2)
Cancer type, n (%)				
HL[g]	120 (10.1)	102 (11.4)	46 (10.1)	27 (11.9)
NHL-HG[h]	676 (56.7)	484 (54.3)	260 (57.0)	125 (55.1)
NHL-LG[i]	280 (23.5)	224 (25.1)	114 (25.0)	56 (24.7)
CLL[j]	116 (9.7)	82 (9.2)	36 (7.9)	19 (8.4)
Primary treatment, n (%)				
Active surveillance	315 (26.4)	252 (28.3)	113 (24.8)	53 (23.3)
CT[k] alone	522 (43.8)	405 (45.4)	222 (48.7)	109 (48.0)
RT[l] alone	80 (6.7)	70 (7.8)	31 (6.8)	11 (4.8)
CT + RT	103 (8.6)	90 (10.1)	51 (11.2)	29 (12.8)
SCT[m]/CT/ RT	30 (2.5)	25 (2.8)	20 (4.4)	16 (7.0)
Other	56 (4.7)	39 (4.4)	16 (3.5)	7 (3.1)
Unknown	87 (7.3)	11 (1.2)	3 (0.7)	2 (0.9)
Number of self-reported comorbidities, mean (SD)	N/A	N/A	1.3 (1.2)	1.1 (1.1)
Personality traits, mean (SD)				
Openness	N/A	N/A	3.4 (0.6)	3.5 (0.6)
Conscientiousness	N/A	N/A	3.7 (0.5)	3.7 (0.4)
Extraversion	N/A	N/A	3.5 (0.6)	3.5 (0.6)
Agreeableness	N/A	N/A	3.8 (0.5)	3.8 (0.4)
Neuroticism	N/A	N/A	2.5 (0.7)	2.4 (0.6)
Information preferences, n (%)				
All available information	N/A	N/A	211 (46.3)	126 (55.5)
Only positive information	N/A	N/A	65 (9.9)	23 (10.1)
Limited information	N/A	N/A	170 (37.3)	78 (34.4)
Internet use (yes)	N/A	N/A	370 (81.1)	220 (96.9)

[a]See Figure 1 for the flowchart of patient selection. The groups in columns are not mutually exclusive; RCT participants are also included in the groups of the other 3 columns.

[b]RCT: randomized controlled trial.

[c]Low education level: none/primary school.

[d]N/A: data not available.

[e]Medium education level: lower general secondary education/vocational training.

[f]High education level: preuniversity education/high-level vocational training/university.

[g]HL: Hodgkin lymphoma.

[h]NHL-HG: high-grade nonHodgkin lymphoma.

[i]NHL-LG: low-grade nonHodgkin lymphoma.

[j]CLL: chronic lymphocytic leukemia.

[k]CT: chemotherapy.

[l]RT: radiotherapy.

[m]SCT: stem cell transplantation.

Patients Invited to Participate

A flowchart of the patient selection process is shown in Figure 1. Of all selected patients, 24.81% (296/1193) were excluded after verifying eligibility by their treating hematologists for the following reasons: deceased, severe psychopathology, too ill, insufficient proficiency of the Dutch language, and cognitive impairment. In addition, 156 patients were not eligible for other reasons, including 36 patients who received treatment or follow up in another hospital and 20 patients for whom the ultimate diagnosis did not meet our inclusion criteria (eg, myelodysplastic syndrome, acute lymphoblastic leukemia). The remaining 100 patients were excluded by the hematologists for unknown reasons. Furthermore, 5 patients were excluded as they declined participation in previous studies within the PROFILES registry. After exclusion of these patients, 892 patients (74.77%) were invited to participate and completed a questionnaire. Patients invited to participate did not significantly differ from all patients selected from the NCR in terms of age (P=.38) and sex (P=.07) (Table 1).

Respondents

Among the 892 invited patients, 456 patients (51.1%) responded and completed either a web-based or paper questionnaire. The mean age of all respondents (Table 1) was comparable with that of nonrespondents (63.8 years, P=.43), and the majority of the respondents were also men. Respondents did not differ from all patients selected from the NCR in terms of age (P=.81) and sex (P=.26). Respondents were more often actively treated than nonrespondents (75% vs 66%, P=.01). Half of the respondents (229/456, 50.2%) completed a paper questionnaire, whereas the other half (227/456, 49.8%) completed a web-based questionnaire and were enrolled in the RCT. Nearly half of the respondents stated that they would like to receive all available information, with a lower proportion preferring limited information, and even less indicating that they would like to receive only positive information about the illness (Table 1). Approximately 82% of all respondents reported using the internet.

Randomized Controlled Trial Participants

A quarter of all invited patients (227/892) participated by completing a web-based questionnaire, which resulted in a participation rate of the RCT of 25.4%. The mean age of RCT participants was slightly lower than that of nonparticipants (ie, nonrespondents and paper respondents (n=665, 65.3 years, P<.001) with a slightly higher mean time since diagnosis, and comprised a higher proportion of men (70.9%, 161/227 vs 57.0%, 379/665, P<.001). The proportion of patients who were actively treated were comparable between RCT participants and nonparticipants (75.8%, 172/227 vs 69.0%, 459/665, P=.13). In addition, RCT participants were significantly younger than respondents, patients invited to participate, and patients selected from the NCR (all P<.001). Furthermore, there was a higher proportion of men among RCT participants compared with all patients invited to participate and all patients selected from the NCR (both P<.001).

Randomized Controlled Trial Participants Versus Paper Participants

RCT participants were younger than paper respondents (60.7 vs 68.3 years, P<.001). In addition, RCT participants were more often male (71% vs 57%, P=.002), more highly educated (50% vs 20%, P<.001), and more often had a partner (84% vs 75%, P=.02). No significant differences were found between RCT and paper respondents regarding cancer type or primary treatment (P=.54 and P=.06, respectively). RCT participants also reported fewer comorbidities than paper respondents (1.1 vs 1.4, P=.02).

Concerning personality traits, RCT participants had lower scores on neuroticism (2.4 vs 2.6, P=.003) and higher scores on openness to experience (3.5 vs 3.4, P=.002) than paper respondents, although effect sizes were small (Cohen d=0.29 and 0.28, respectively). With respect to information preferences, the majority of RCT participants stated a preference for receiving all available information, whereas only 38.9% (89/229) of paper respondents indicated this preference (P=.001). Conversely, paper respondents more often preferred receiving limited information (41.9%, 96/227 vs 34.4%, 78/227; P=.001). Furthermore, RCT participants more often used the internet (96.9%, 220/227 vs 66.4%, 152/229; P<.001) (Table 1).

Emotional distress, as measured with the heiQ, was significantly lower among RCT participants compared with the score for paper respondents (Table 2), although the effect size was small (Cohen d=0.25). No significant differences were observed regarding other self-management skills between the RCT and paper groups. In addition, no significant differences were observed in the proportion of patients with psychological distress between RCT participants and paper respondents, although paper participants seemed to have higher mean scores. Furthermore, no differences were observed between RCT

participants and paper respondents regarding satisfaction with overall information provision (Table 2).

Table 2. Differences in randomized controlled trial (RCT) outcome variables at baseline between RCT participants and paper respondents.

Outcome variable	RCT participants (N=227)	Paper respondents (N=229)	*P* value	Cohen *d*
Psychological distress (yes), n (%)	34 (15.0)	45 (19.7)	.18	
Psychological distress[a], mean (SD)	6.5 (5.9)	7.5 (6.1)	.06	0.18
Self-management skills[b], mean (SD)				
Health-directed behavior	3.3 (0.6)	3.2 (0.6)	.12	0.14
Positive and active engagement in life	3.2 (0.5)	3.1 (0.5)	.05	0.18
Self-monitoring and insight	3.0 (0.4)	3.1 (0.4)	.62	0.05
Constructive attitudes and approaches	3.3 (0.5)	3.3 (0.5)	.66	0.04
Skill and technique acquisition	2.9 (0.5)	3.0 (0.5)	.33	0.09
Social integration and support	3.2 (0.5)	3.2 (0.5)	.77	0.03
Health services navigation	3.3 (0.4)	3.3 (0.4)	.54	0.06
Emotional distress	1.8 (0.5)	1.9 (0.6)	**.**01	0.25
Satisfaction with overall information provision, n (%)			.29	
Very unsatisfied	1 (0.4)	2 (0.9)		
Unsatisfied	9 (4.0)	8 (3.5)		
Neither	49 (21.6)	41 (17.9)		
Satisfied	126 (55.5)	143 (62.4)		
Very satisfied	41 (18.1)	27 (11.8)		

[a]Scale 0-42; a higher score indicates more psychological distress.

[b]Scale 1-4; higher scores indicate better status or self-management, except for emotional distress, in which higher scores indicate higher distress.

Discussion

Principal Findings

This reach analysis among RCT participants within a population-based sample showed a selective reach with an underrepresentation of older patients, women, and those with a medium to low level of education. In addition, our RCT participants may represent individuals with relatively better psychological well-being as scores for emotional distress were lower in this group.

Approximately a quarter of the population-based sample of patients with lymphoma and CLL were assessed to be not eligible for the study for various reasons (eg, deceased, severe psychopathology or cognitive impairment, a different diagnosis was ultimately made). Among the eligible patients who were invited to participate, 51% responded and completed a questionnaire, half of whom completed the web-based questionnaire and were enrolled in the RCT, resulting in a participation rate of 25%. This means that only one in four of all eligible patients actually participated in the RCT. This participation rate was lower compared with that of an RCT on the fully automated electronic health (eHealth) application Oncokompas that supports cancer survivors in their self-management (48%) [24]. Similar to the patients in our RCT, patients in the Oncokompas RCT were selected from the population-based NCR. However, cancer survivors in the Oncokompas RCT were first invited in an online survey study on supportive care and eHealth to assess internet use. Their participation rate was calculated as the number of RCT participants divided by the number of eligible respondents of the survey (access to the internet and email address). Thus, their group of eligible respondents was more selective compared with our sample. In our sample, only 82% of all respondents used the internet, and this percentage may be even lower among all patients invited to participate.

The results of the current study demonstrate that the RCT participants were younger, more often men, and more often actively treated compared to nonparticipants. Thus, the sample of RCT participants may not be representative of the target population. Therefore, even though the sample size reached the required number of patients [12], this sample may not be reliable for drawing conclusions about the target population. Furthermore, the effects of the intervention on the target population may be different from the effects that were found in the RCT sample [10,25].

These results also provide information about the response rate of observational research, which was 51%. This is comparable with response rates from other population-based studies on quality of life among lymphoma survivors in Germany (54.7%) [26] and the United States (54.8%) [27]. However, the current response rate is lower compared with that reported from earlier observational research within our study group at approximately

80%, despite similar patients and recruitment procedures [28]. This might be explained by the knowledge that the more information that is disclosed about the study—which is inherently more for an RCT than for observational research—the higher the proportion of nonrespondents [29]. Patients received abundant information along with the invitation, especially about participation in an RCT and randomization. The amount of information, as well as the knowledge of being randomized when completing a web-based questionnaire, may have deterred patients from participating. In addition, the type of intervention may have influenced the participation rate, as the majority of patients did not have problems with emotional adjustment to having cancer and therefore may have been less interested in a self-management intervention. Another explanation may be related to the fact that participation and response rates for health-related research have been declining over the past several years [30,31], and potential participants are faced with an increasing number of requests to participate in studies. This may result in patients refusing to participate in all studies [32].

We further compared the characteristics of RCT participants with those of paper respondents. RCT participants, who completed the web-based questionnaire, were younger, more often men, and more highly educated than the paper respondents, which are similar to the characteristics from previous observational studies within our study group [33,34]. Highly educated patients more often display prosocial behavior than patients with lower levels of education, and therefore the former group may be more likely to participate in an RCT for altruistic reasons [35]. In addition to differences in sociodemographic characteristics, RCT participants reported lower scores related to neuroticism and higher scores related to openness than paper respondents. In addition, information preferences slightly differed between RCT and paper respondents, as RCT participants more often wanted to receive all available information on their disease. RCT participants also more often reported using the internet. To complete a web-based questionnaire, and subsequently be enrolled in a web-based self-management intervention RCT, patients must not only be able to use a computer but also be sufficiently skilled in browsing the internet [33]. Although there seems to be a trend of older individuals becoming more active online [33,36], there is still a subgroup of patients who do not use the internet and thus have no access to a web-based questionnaire or internet-based intervention.

Despite these various differences between RCT participants and paper participants, baseline scores on self-management skills, satisfaction with information provision, and psychological distress appeared to be comparable between these groups, although scores for emotional distress were slightly lower among RCT participants.

Strengths

The strengths of this study include its unique setting. As patients were recruited from the population-based NCR, we had information on sociodemographic and clinical characteristics of both RCT participants and nonparticipants. In addition, as the RCT was embedded in the PROFILES registry, we were able to assess differences between RCT participants and paper respondents on sociodemographic and clinical characteristics, in addition to personality traits, information preferences, and baseline psychological distress, self-management skills, and satisfaction with information provision (ie, RCT outcome variables). This information provided the opportunity to determine both the reach and generalizability of the RCT sample.

Limitations

The current study has some limitations. Although information regarding sociodemographic and clinical characteristics of nonparticipants was available, we did not have information about nonparticipants' reasons for declining participation or their physical and psychological health. Therefore, it remains unclear whether the physical and psychological health of RCT participants is similar to that of nonparticipants. In a previous study that assessed the generalizability of the results of observational research among cancer survivors by comparing characteristics of participants and nonparticipants, sensitivity analysis demonstrated that quality of life might be lower among nonparticipants [34]. As RCT participants may have a systematically higher quality of life or report fewer symptoms compared with nonparticipants, observed outcomes may represent a group of healthier patients with better outcomes. This may lead to circumspection in generalizing the results of an RCT to the target population. It is important to keep this in mind when interpreting RCT results that may only represent a minority of the target population.

Conclusions

The participation rate in the RCT was 25%. RCT participants may be not representative of the target population owing to different sociodemographic and clinical characteristics. RCT results should be considered with caution, as RCT participants represent a minority of the target population, and may actually be those least in need of the intervention.

Acknowledgments

This study was financially supported by the Jonker-Driessen Foundation. The study funder had no role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication.

Authors' Contributions

LPJA was responsible for patient recruitment, data collection, data analysis, and drafting the manuscript. SO was responsible for data collection and was a major contributor to drafting the manuscript. EFMP, DEI, MO, RvdG, and MRN were responsible for

data collection and contributed to writing the manuscript. LVvdP-F was project leader and was a major contributor in drafting the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Leykin Y, Thekdi SM, Shumay DM, Muñoz RF, Riba M, Dunn LB. Internet interventions for improving psychological well-being in psycho-oncology: review and recommendations. Psychooncology 2012 Sep;21(9):1016-1025 [FREE Full text] [doi: 10.1002/pon.1993] [Medline: 21608075]
2. Hariton E, Locascio JJ. Randomised controlled trials - the gold standard for effectiveness research. BJOG 2018 Dec;125(13):1716. [doi: 10.1111/1471-0528.15199] [Medline: 29916205]
3. Kim SH, Kim K, Mayer DK. Self-Management Intervention for Adult Cancer Survivors After Treatment: A Systematic Review and Meta-Analysis. Oncol Nurs Forum 2017 Nov 01;44(6):719-728. [doi: 10.1188/17.ONF.719-728] [Medline: 29052663]
4. van Heuvelen MJG, Hochstenbach JBM, Brouwer WH, de Greef MHG, Zijlstra GAR, van Jaarsveld E, et al. Differences between participants and non-participants in an RCT on physical activity and psychological interventions for older persons. Aging Clin Exp Res 2005 Jun;17(3):236-245. [doi: 10.1007/bf03324603] [Medline: 16110738]
5. Donkin L, Hickie IB, Christensen H, Naismith SL, Neal B, Cockayne NL, et al. Sampling bias in an internet treatment trial for depression. Transl Psychiatry 2012 Oct 23;2:e174 [FREE Full text] [doi: 10.1038/tp.2012.100] [Medline: 23092978]
6. Kinder BW, Sherman AC, Young LR, Hagaman JT, Oprescu N, Byrnes S, et al. Predictors for clinical trial participation in the rare lung disease lymphangioleiomyomatosis. Respir Med 2010 Apr;104(4):578-583 [FREE Full text] [doi: 10.1016/j.rmed.2009.09.023] [Medline: 19962873]
7. Stuart EA, Bradshaw CP, Leaf PJ. Assessing the generalizability of randomized trial results to target populations. Prev Sci 2015 Apr;16(3):475-485 [FREE Full text] [doi: 10.1007/s11121-014-0513-z] [Medline: 25307417]
8. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999 Sep;89(9):1322-1327. [doi: 10.2105/ajph.89.9.1322] [Medline: 10474547]
9. Glasgow RE, McCaul KD, Fisher KJ. Participation in worksite health promotion: a critique of the literature and recommendations for future practice. Health Educ Q 1993;20(3):391-408. [doi: 10.1177/109019819302000309] [Medline: 8307762]
10. Martínez-Mesa J, González-Chica DA, Bastos JL, Bonamigo RR, Duquia RP. Sample size: how many participants do I need in my research? An Bras Dermatol 2014;89(4):609-615 [FREE Full text] [doi: 10.1590/abd1806-4841.20143705] [Medline: 25054748]
11. Martínez-Mesa J, González-Chica DA, Duquia RP, Bonamigo RR, Bastos JL. Sampling: how to select participants in my research study? An Bras Dermatol 2016;91(3):326-330 [FREE Full text] [doi: 10.1590/abd1806-4841.20165254] [Medline: 27438200]
12. Arts LPJ, van de Poll-Franse LV, van den Berg SW, Prins JB, Husson O, Mols F, et al. Lymphoma InterVEntion (LIVE) - patient-reported outcome feedback and a web-based self-management intervention for patients with lymphoma: study protocol for a randomised controlled trial. Trials 2017 Apr 28;18(1):199 [FREE Full text] [doi: 10.1186/s13063-017-1943-2] [Medline: 28454583]
13. van de Poll-Franse LV, Horevoorts N, van Eenbergen M, Denollet J, Roukema JA, Aaronson NK, Profiles Registry Group. The Patient Reported Outcomes Following Initial treatment and Long term Evaluation of Survivorship registry: scope, rationale and design of an infrastructure for the study of physical and psychosocial outcomes in cancer survivorship cohorts. Eur J Cancer 2011 Sep;47(14):2188-2194 [FREE Full text] [doi: 10.1016/j.ejca.2011.04.034] [Medline: 21621408]
14. Percy C, Fritz A, Jack A, Shanmugarathan S, Sobin L, Parkin D, et al. International Classification Of Diseases For Oncology 3rd edition (ICD-O-3). Geneva: World Health Organization; 2000.
15. Sangha O, Stucki G, Liang MH, Fossel AH, Katz JN. The Self-Administered Comorbidity Questionnaire: a new method to assess comorbidity for clinical and health services research. Arthritis Rheum 2003 Apr 15;49(2):156-163. [doi: 10.1002/art.10993] [Medline: 12687505]
16. John O, Donahue E, Kentle R. The Big Five Inventory--Versions 4a and 54. Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research; 1991.
17. Hayes N, Joseph S. Big 5 correlates of three measures of subjective well-being. Pers Ind Diff 2003 Mar;34(4):723-727. [doi: 10.1016/s0191-8869(02)00057-0]
18. McCrae R, Costa P. Introduction to the empirical and theoretical status of the five-factor model of personality traits. In: Widiger TA, Costa PTJ, editors. Personality disorders and the five-factor model of personality. Washington, DC: American Psychological Association; 2013:15-27.
19. Thomas R, Kaminski E, Stanton E, Williams M. Measuring information strategies in oncology - developing an information satisfaction questionnaire. Eur J Cancer Care 2004 Mar;13(1):65-70. [doi: 10.1111/j.1365-2354.2004.00445.x] [Medline: 14961777]

20. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
21. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. J Psychosom Res 2002 Feb;52(2):69-77. [doi: 10.1016/s0022-3999(01)00296-3] [Medline: 11832252]
22. Singer S, Kuhnt S, Götze H, Hauss J, Hinz A, Liebmann A, et al. Hospital anxiety and depression scale cutoff scores for cancer patients in acute care. Br J Cancer 2009 Mar 24;100(6):908-912. [doi: 10.1038/sj.bjc.6604952] [Medline: 19240713]
23. Osborne RH, Elsworth GR, Whitfield K. The Health Education Impact Questionnaire (heiQ): an outcomes and evaluation measure for patient education and self-management interventions for people with chronic conditions. Patient Educ Couns 2007 May;66(2):192-201. [doi: 10.1016/j.pec.2006.12.002] [Medline: 17320338]
24. van der Hout A, van Uden-Kraan CF, Holtmaat K, Jansen F, Lissenberg-Witte BI, Nieuwenhuijzen GAP, et al. Role of eHealth application Oncokompas in supporting self-management of symptoms and health-related quality of life in cancer survivors: a randomised, controlled trial. Lancet Oncol 2020 Jan;21(1):80-94. [doi: 10.1016/s1470-2045(19)30675-8]
25. Martinson BC, Crain AL, Sherwood NE, Hayes MG, Pronk NP, O'Connor PJ. Population reach and recruitment bias in a maintenance RCT in physically active older adults. J Phys Act Health 2010 Jan;7(1):127-135 [FREE Full text] [doi: 10.1123/jpah.7.1.127] [Medline: 20231764]
26. Hammersen F, Lewin P, Gebauer J, Kreitschmann-Andermahr I, Brabant G, Katalinic A, et al. Sleep quality and health-related quality of life among long-term survivors of (non-) Hodgkin lymphoma in Germany. PLoS One 2017;12(11):e0187673 [FREE Full text] [doi: 10.1371/journal.pone.0187673] [Medline: 29107959]
27. Arora NK, Hamilton AS, Potosky AL, Rowland JH, Aziz NM, Bellizzi KM, et al. Population-based survivorship research using cancer registries: a study of non-Hodgkin's lymphoma survivors. J Cancer Surviv 2007 Mar;1(1):49-63. [doi: 10.1007/s11764-007-0004-3] [Medline: 18648945]
28. Oerlemans S, Issa DE, van den Broek EC, Nijziel MR, Coebergh JWW, Huijgens PC, et al. Health-related quality of life and persistent symptoms in relation to (R-)CHOP14, (R-)CHOP21, and other therapies among patients with diffuse large B-cell lymphoma: results of the population-based PHAROS-registry. Ann Hematol 2014 Oct;93(10):1705-1715. [doi: 10.1007/s00277-014-2099-8] [Medline: 24807105]
29. Simes RJ, Tattersall MH, Coates AS, Raghavan D, Solomon HJ, Smartt H. Randomised comparison of procedures for obtaining informed consent in clinical trials of treatment for cancer. Br Med J (Clin Res Ed) 1986 Oct 25;293(6554):1065-1068 [FREE Full text] [doi: 10.1136/bmj.293.6554.1065] [Medline: 3094776]
30. Keiding N, Louis TA. Perils and potentials of self-selected entry to epidemiological studies and surveys. JR Stat Soc A 2016 Jan 18;179(2):319-376. [doi: 10.1111/rssa.12136]
31. Drivsholm T, Eplov LF, Davidsen M, Jørgensen T, Ibsen H, Hollnagel H, et al. Representativeness in population-based studies: a detailed description of non-response in a Danish cohort study. Scand J Public Health 2006;34(6):623-631. [doi: 10.1080/14034940600607616] [Medline: 17132596]
32. Galea S, Tracy M. Participation rates in epidemiologic studies. Ann Epidemiol 2007 Sep;17(9):643-653. [doi: 10.1016/j.annepidem.2007.03.013] [Medline: 17553702]
33. Horevoorts NJ, Vissers PA, Mols F, Thong MS, van de Poll-Franse LV. Response rates for patient-reported outcomes using web-based versus paper questionnaires: comparison of two invitational methods in older colorectal cancer patients. J Med Internet Res 2015;17(5):e111 [FREE Full text] [doi: 10.2196/jmir.3741] [Medline: 25953059]
34. de Rooij BH, Ezendam NPM, Mols F, Vissers PAJ, Thong MSY, Vlooswijk CCP, et al. Cancer survivors not participating in observational patient-reported outcome studies have a lower survival compared to participants: the population-based PROFILES registry. Qual Life Res 2018 Dec;27(12):3313-3324 [FREE Full text] [doi: 10.1007/s11136-018-1979-0] [Medline: 30167937]
35. Schwartz CE, Fox BH. Who says yes? Identifying selection biases in a psychosocial intervention study of multiple sclerosis. Soc Sci Med 1995 Feb;40(3):359-370. [doi: 10.1016/0277-9536(94)e0092-7] [Medline: 7899948]
36. van Eenbergen MCHJ, Vromans RD, Boll D, Kil PJM, Vos CM, Krahmer EJ, et al. Changes in internet use and wishes of cancer survivors: A comparison between 2005 and 2017. Cancer 2020 Jan 15;126(2):408-415 [FREE Full text] [doi: 10.1002/cncr.32524] [Medline: 31580497]

Abbreviations

CLL: chronic lymphocytic leukemia
eHealth: electronic health
HeiQ: Health Education Impact Questionnaire
LIVE: Lymphoma InterVEntion
NCR: Netherlands Cancer Registry
PROFILES: Patient Reported Outcomes Following Initial treatment and Long term Evaluation of Survivorship
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 12.11.19; peer-reviewed by L Russell, F Sartor; comments to author 27.12.19; revised version received 30.01.20; accepted 01.02.20; published 14.05.20.

Please cite as:
Arts LPJ, Oerlemans S, Posthuma EFM, Issa DE, Oosterveld M, van der Griend R, Nijziel MR, van de Poll-Franse LV
Web-Based Self-Management for Patients With Lymphoma: Assessment of the Reach of Intervention of a Randomized Controlled Trial
J Med Internet Res 2020;22(5):e17018
URL: http://www.jmir.org/2020/5/e17018/
doi:10.2196/17018
PMID:32406858

Original Paper

A Mobile App Lifestyle Intervention to Improve Healthy Nutrition in Women Before and During Early Pregnancy: Single-Center Randomized Controlled Trial

Matthijs R van Dijk[1], MD, PhD; Maria P H Koster[1], MD, PhD; Elsje C Oostingh[1], MD; Sten P Willemsen[2], PhD; Eric A P Steegers[1], MD, PhD; Régine P M Steegers-Theunissen[1], MD, PhD

[1]Department of Obstetrics and Gynaecology, Erasmus Medical Center, University Medical Center Rotterdam, Rotterdam, Netherlands

[2]Department of Biostatistics, Erasmus Medical Center, University Medical Center Rotterdam, Rotterdam, Netherlands

Corresponding Author:
Régine P M Steegers-Theunissen, MD, PhD
Department of Obstetrics and Gynaecology
Erasmus Medical Center
University Medical Center Rotterdam
PO Box 2040
Rotterdam
Netherlands
Phone: 31 107038254
Email: r.steegers@erasmusmc.nl

Abstract

Background: Unhealthy nutrition contributes to the worldwide rising prevalence of noncommunicable diseases. As most adverse reproductive outcomes originate during the periconception period, effective interventions targeting this period are needed. Therefore, we developed the lifestyle intervention Smarter Pregnancy to empower women to adapt a healthy diet prior to conception and during early pregnancy and performed a randomized controlled trial.

Objective: The objectives of this trial were to investigate compliance and effectiveness in women using the Smarter Pregnancy program.

Methods: Women aged between 18 and 45 years who were contemplating pregnancy or <13 weeks pregnant and their male partners living in the urban area of Rotterdam, the Netherlands, were eligible for participation. After baseline screening, the intervention group received personal online coaching based on identified inadequate intakes of vegetables, fruits, and folic acid supplements. The sum of these risk factors was used as a dietary risk score (DRS), ranging from 0 (healthy) to 9 (unhealthy). The control group did not receive coaching. We applied an intention-to-treat principle and used a multivariable linear regression model to evaluate the change in DRS after 24 weeks. Compliance was defined as the percentage of women who completed the screening questionnaire at 24 weeks.

Results: Of women recruited, 81.2% (177/218) completed the program (intervention: 91/218, 83.5%; control: 86/218, 78.9%; *P*=.95). After 24 weeks, the reduction in DRS of women in the intervention group was significantly larger than in the control group (β=.75, 95% CI 0.18-1.34). This reduction was mainly due to increased vegetable intake (β=.55, 95% CI 0.25-0.86).

Conclusions: The high compliance and the larger improvements in nutritional behaviors, especially vegetable intake, in women in the intervention group emphasizes the effectiveness of empowering women by using the lifestyle change intervention Smarter Pregnancy.

Trial Registration: Netherlands Trial Register: NL3927; https://www.trialregister.nl/trial/3927

International Registered Report Identifier (IRRID): RR2-10.1186/s12884-017-1228-5

(J Med Internet Res 2020;22(5):e15773) doi:10.2196/15773

KEYWORDS

mHealth; preconception care; nutrition; pregnancy

Introduction

Unhealthy nutrition contributes to the development of noncommunicable diseases (NCDs) such as obesity, diabetes, and cardiovascular and metabolic disease [1-5]. In recent decades, the worldwide prevalence of NCDs and corresponding mortality rates have increased rapidly [6]. Vitamin deficiencies and high caloric intake combined with inadequate physical exercise are key risk factors for metabolic and endocrine derangements that contribute to obesity and a wide spectrum of NCDs [7-9]. These risk factors are also highly prevalent in women and men during the reproductive phase of their life, with significant consequences for fertility, growth, and development of the offspring [10-17]. Moreover, unhealthy nutrition and lifestyle also confer increased transgenerational risks for offspring in developing NCDs in later life [2,5,18-20].

There is increasing evidence for a need for effective interventions to improve nutrition and other modifiable risk factors in women who are contemplating pregnancy, particularly in the periconception period (ie, the period 14 weeks prior to conception up to 10 weeks after conception) [19,21]. As most adverse reproductive and pregnancy outcomes originate during this period, it is considered the earliest window of opportunity for interventions. However, since the periconception period is often neglected in regular health care, with specific periconception care rarely implemented, the prevalence of these modifiable risk factors still remains very high in the population of reproductive age [10,22].

In order to translate the scientific evidence currently available into accessible periconception care, various barriers need to be overcome. These barriers include the lack of intrinsic motivation for changing lifestyle in the target population; low levels of awareness; and a lack of clarity regarding responsibility, organization, and costs [23-26]. One way of overcoming some of these barriers is to make use of recent developments in electronic health (eHealth). These include using the broad range of functions available on mobile phones and handheld devices, with or without internet access, also known as mobile health (mHealth) [27,28]. Indeed, the global use of smartphones has opened new doors for health care delivery: new and innovative approaches in the fields of preventive and personalized medicine can provide patients with both general information and individualized content [27,29,30]. In 2011, we launched a lifestyle change intervention called Smarter Pregnancy that aims to empower women and men to adopt healthy nutrition and lifestyle behaviors before and during pregnancy. This program is based on models for behavior change and existing evidence regarding the impact of nutrition and lifestyle on fertility and maternal pregnancy and birth outcomes and provides individual coaching on five major risk factors: inadequate vegetable, fruit, and folic acid supplement intake and smoking and alcohol consumption [10,12,13,28,31-34]. Since an inadequate daily intake of fruit and vegetables are the most prevalent risk factors for unhealthy nutrition and users appreciate interventions that are as simple as possible, we hypothesized that stimulating the intakes of these healthy food groups would result in a more balanced and healthy diet and lifestyle in general. We designed a randomized controlled trial to determine compliance with the Smarter Pregnancy intervention and investigate whether use of the program empowers women to improve nutrition prior to conception and during early pregnancy.

Methods

Trial Design, Participants, and Recruitment

A detailed study protocol is published elsewhere [34]. In short, women between aged 18 and 45 years were considered eligible for inclusion in this study if they were in possession of a smartphone with internet access, resided in the Netherlands, and were contemplating pregnancy or already pregnant (<13 weeks of pregnancy). We excluded women if they had insufficient knowledge or understanding of the Dutch language, if they were being treated by a dietician to lose weight in the context of fertility treatment, or if they were on a vegan diet. Body mass index (BMI) was not an exclusion criterion. Dutch-speaking male partners with smartphones were also invited to participate unless they were receiving dietary advice or were on a vegan diet.

We performed a single-center, open randomized controlled trial in the urban area of Rotterdam, the Netherlands. There was no blinding of participants, involved health care professionals, or involved researchers. Women eligible for inclusion were invited to participate by a health care professional working in one of the following locations in this area: one academic hospital, four teaching hospitals, four midwifery practices, and several children's daycare and child health centers. After online registration, each woman was contacted by a researcher to verify her eligibility, provide her with more details about the study, answer any questions about the Smarter Pregnancy program, and confirm her inclusion in the study. After the researcher had verified the eligibility of women willing to participate in the study, inclusion occurred by signing an online (ie, digital) patient informed consent form, which was sent by email to the participant through a secure study email account. Participants were asked to print and sign the informed consent form to ensure compliance with the guidelines laid down in the Declaration of Helsinki. Participants were able to resign from the study at any time without having to give a reason. All procedures involving patients were approved by the medical ethical and institutional review board of the Erasmus Medical Center, University Medical Centre, Rotterdam, the Netherlands. Trial was registered at Netherlands Trial Register [NL3927].

Randomization

Randomization was stratified according to the location from which the participants had been recruited. A preprogrammed permuted blocking design (two intervention and two control allocations per block) ensured that the number of women from the different locations was balanced between the two treatment groups and allocation into groups was concealed from the researchers. Men were always assigned the same group as their female partner.

Intervention

The design and development of the lifestyle change intervention led to the availability of two versions that we could use in our study: a full version that included all functionality and

personalized interaction (the intervention), and a modified version that had limited functionality and no personalized interaction, which was used in the control group. Detailed information on the program can be found in the study protocol [34]. Because the focus of this study was on evaluating the change in the intakes of vegetables and fruits, details on the intake of other food groups and cessation of smoking and alcohol consumption are not further addressed.

Men and women in the intervention group received tailored coaching based on their answers of the baseline questionnaire to questions regarding vegetable, fruit, and folic acid supplement intake. Vegetable intake and fruit intake were both subdivided into a risk score of 0, 1.5, or 3, where 0 represented an adequate daily intake (vegetable intake of ≥200 grams per day or a fruit intake of ≥2 pieces per day). A score of 1.5 represented a nearly adequate intake (vegetable intake of 150 to 200 grams per day or a fruit intake of 1.5 to 2 pieces per day). A score of 3 represented an inadequate daily intake (vegetable intake <150 grams per day or a fruit intake of <1.5 pieces per day). Folic acid supplement use was considered adequate (score 0) or inadequate (score 3) based on the international recommended dose of 400 µg per day. The dietary risk score (DRS) was calculated as the sum of the scores for vegetable, fruit, and folic acid supplement intake, thus ranging from 0 to 9 in women, in which 9 was the most unhealthy risk score. In men, the DRS ranged from 0 to 6, as they did not receive any coaching regarding folic acid supplement use.

The tailored coaching comprised a maximum of three emails or text messages per week. These emails and messages contained seasonal recipes, incentives, feedback, recommendations, and additional questions regarding the participant's diet. Progress regarding the adoption of healthy behavior was monitored using online questionnaires at 6, 12, 18, and 24 weeks in the intervention group, while the control group only received these questionnaires at 12 and 24 weeks.

The first follow-up study questionnaire was sent at 36 weeks (ie, 12 weeks after the final screening questionnaire) and contained the same questions on nutrition, lifestyle, and pregnancy status as the other online questionnaires at baseline, 12, and 24 weeks.

All participants were given access to a personal online webpage that provided access to additional modules (ie, apps) that promoted physical activity, a calendar to improve compliance with hospital appointments and taking their folic acid supplements, and a module to monitor the safety of any prescribed medication.

Outcome Measures

The main outcome measures of this study were compliance of all participants, defined as the percentage of participants who completed the online screening at 24 weeks, and degree of improvement in nutrition in women 24 weeks after starting the Smarter Pregnancy program, as reflected by a reduction in the DRS.

Statistical Analysis

We analyzed data from all participants, those who completed the Smarter Pregnancy program and those who resigned prematurely, whereby missing data were handled using the last-observation-carried-forward method.

For all participants, the DRS was calculated at baseline (t=0), 12 weeks (t=12), and 24 (t=24) weeks. In further analyses, we included all women with a DRS>0 at baseline since these women were able to improve unhealthy behaviors and thereby reduce their risk scores. The primary analysis was based on intention to treat. The difference in differences principle was used to analyze the continuous outcome measures used in a multivariable linear regression model, adjusted for the baseline value of the DRS. Bootstrapping was performed on all analyses because residuals of the linear regression analyses were not normally distributed [35]. All analyses were performed using SPSS Statistics for Windows version 21.0 (IBM Corp). The analyzed dataset of this trial will be available from the corresponding author upon reasonable request.

Patient Involvement

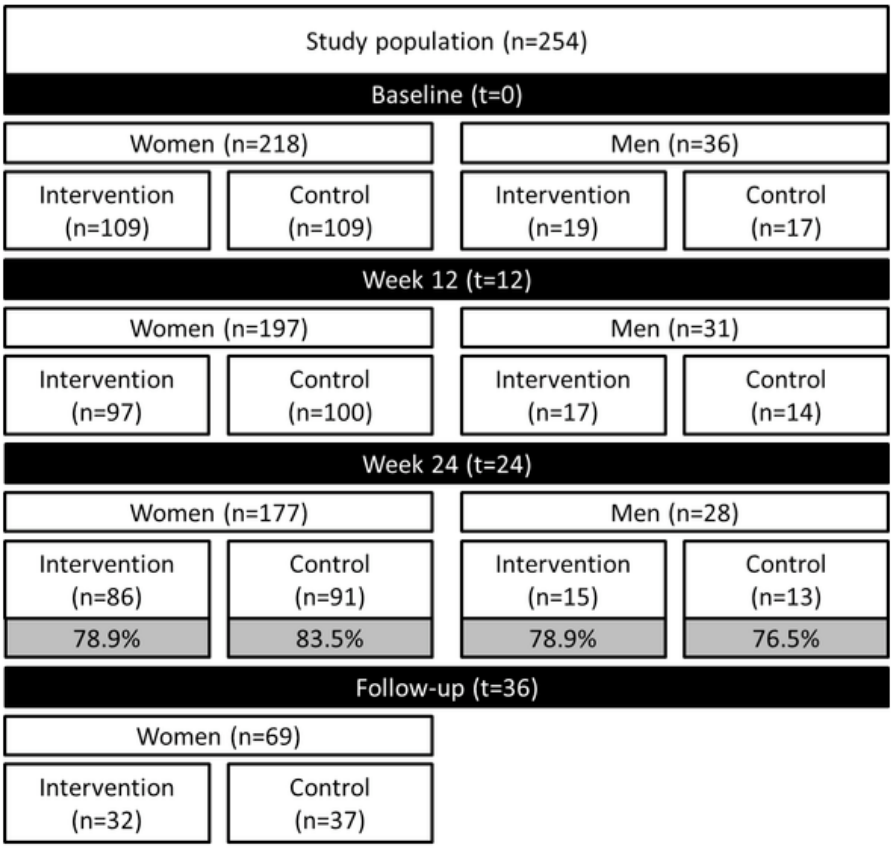
The design of this trial was based partly on patient evaluations obtained during a survey of the Smarter Pregnancy program [12]. During this study, we also received questions and feedback from participants, which we used to optimize trial procedures and improve participant satisfaction.

Results

Participant Characteristics

The study was open from May 2014 until January 2017 and included 218 women and 36 men. After randomization, the intervention group consisted of 109 women and 19 men and the control group of 109 women and 17 men (Figure 1). Baseline characteristics of all women in the study population are shown in Table 1. Median age, median BMI, pregnancy status, and partner participation were similar for women in the intervention and control groups. In our population, the BMI of all participants ranged between 16.3 to 45.9 kg/m^2. In both groups, most women were highly educated and of Dutch origin.

Figure 1. Flowchart of the study population stratified by sex.



Regarding nutrition and lifestyle, in both groups almost two-thirds of women reported an inadequate vegetable intake. Fruit intake was inadequate in about one-third of women in both groups. In both groups, almost 1 in 10 women reported inadequate folic acid supplement use. These figures resulted in a median DRS at baseline of 3 in both groups. The baseline characteristics of the participating men are shown in Multimedia Appendix 1; in both groups, men had a higher DRS and a higher prevalence of smoking compared with women.

Table 1. Baseline characteristics of all women in the intervention and control groups.

Characteristics	Intervention n=109	Control n=109
Age in years, median (IQR[a])	30.6 (5.3)	30.7 (5.7)
Height (cm), median (IQR)	170.0 (9)	170.0 (9)
BMI[b] (kg/m^2), median (IQR)	24.2 (6.0)	23.7 (5.4)
Pregnant at enrollment, n (%)	36 (33.0)	37 (33.9)
Partner participation, n (%)	19 (17.4)	18 (16.5)
Geographic origin, n (%)		
Dutch	83 (76.1)	86 (78.9)
Western	5 (4.6)	2 (1.8)
Nonwestern	15 (13.8)	15 (13.8)
Missing	6 (5.5)	6 (5.5)
Education, n (%)		
High	62 (56.9)	76 (69.7)
Intermediate	37 (33.9)	23 (21.1)
Low	1 (0.9)	3 (2.8)
Missing	9 (8.3)	7 (6.4)
Vegetables, grams per day, n (%)		
<150 (DRS[c] 3)	65 (59.6)	65 (59.6)
150-200 (DRS 1.5)	19 (17.4)	24 (22.0)
≥200 (DRS 0)	25 (22.9)	20 (18.3)
Fruit, pieces per day, n (%)		
<1.5 (DRS 3)	39 (35.8)	37 (33.9)
1.5-2.0 (DRS 1.5)	8 (7.3)	14 (12.8)
≥2.0 (DRS 0)	62 (56.9)	58 (53.2)
Folic acid supplement use, n (%)		
Inadequate (DRS 3)	10 (9.2)	10 (9.2)
Adequate (DRS 0)	99 (90.8)	99 (90.8)
DRS, median (IQR) (DRS 0-9)	3 (4.5)	3 (3.0)
Alcohol consumption, n (%)	85 (78.0)	82 (75.2)
Smoking, n (%)	5 (4.6)	12 (11.0)

[a]IQR: interquartile range.

[b]BMI: body mass index.

[c]DRS: dietary risk score.

Compliance and Dropout

Of women entering the study, compliance was 81.2% (177/208). In the intervention group, compliance was 78.9% (86/109) and in the control group 83.5% (91/109; *P*=.95). In men, overall compliance was 77.8% (28/36), with 78.9% (15/19) in the intervention group and 76.5% (13/17) in the control group (*P*=.59). When we compared the baseline characteristics of all women who completed the program (n=177) with those of women who resigned prematurely (n=41), we observed that women who resigned prematurely had a significantly higher median DRS (*P*=.007) and a significantly lower level of education (*P*=.01; Table 2).

Table 2. Baseline characteristics of all women, stratified by compliance, defined as whether they completed the questionnaire at 24 weeks or resigned before this time point.

Characteristics	Completed n=177	Resigned n=41	*P* value
Age in years, median (IQR[a])	30.8 (6.0)	30.1 (6.0)	.33
Height (cm), median (IQR)	170 (10.0)	169 (8.0)	.86
BMI[b] (kg/m^2), median (IQR)	23.4 (5.8)	25.6 (4.5)	.13
Pregnant at enrollment, n (%)	63 (33.9)	10 (31.3)	.77
Partner participation, n (%)	31 (16.7)	6 (18.8)	.77
Geographic origin, n (%)			.21
Dutch	145 (78.0)	24 (75.0)	
Western	7 (3.8)	0 (0)	
Nonwestern	27 (14.5)	3 (9.4)	
Missing	7 (4.0)	5 (12.2)	
Education, n (%)			.01
High	119 (64.0)	19 (59.4)	
Intermediate	53 (28.5)	7 (21.9)	
Low	3 (1.6)	1 (3.1)	
Missing	11 (6.2)	5 (12.2)	
Vegetables, grams per day, n (%)			.13
<150 (DRS[c] 3)	106 (57.0)	24 (75.0)	
150-200 (DRS 1.5)	38 (20.4)	5 (15.6)	
≥200 (DRS 0)	42 (22.6)	3 (9.4)	
Fruit, pieces per day, n (%)			.06
<1.5 (DRS 3)	59 (31.7)	17 (53.1)	
1.5-2.0 (DRS 1.5)	19 (10.2)	3 (9.4)	
≥2.0 (DRS 0)	108 (58.0)	12 (37.5)	
Folic acid supplement use, n (%)			.48
Inadequate (DRS 3)	16 (8.6)	4 (12.5)	
Adequate (DRS 0)	170 (91.4)	28 (87.5)	
DRS, median (IQR) (DRS 0-9)	3 (4.5)	4.5 (3.0)	.007
Alcohol consumption, n (%)	40 (21.5)	11 (34.4)	.14
Smoking, n (%)	13 (7.0)	4 (12.5)	.28

[a]IQR: interquartile range.

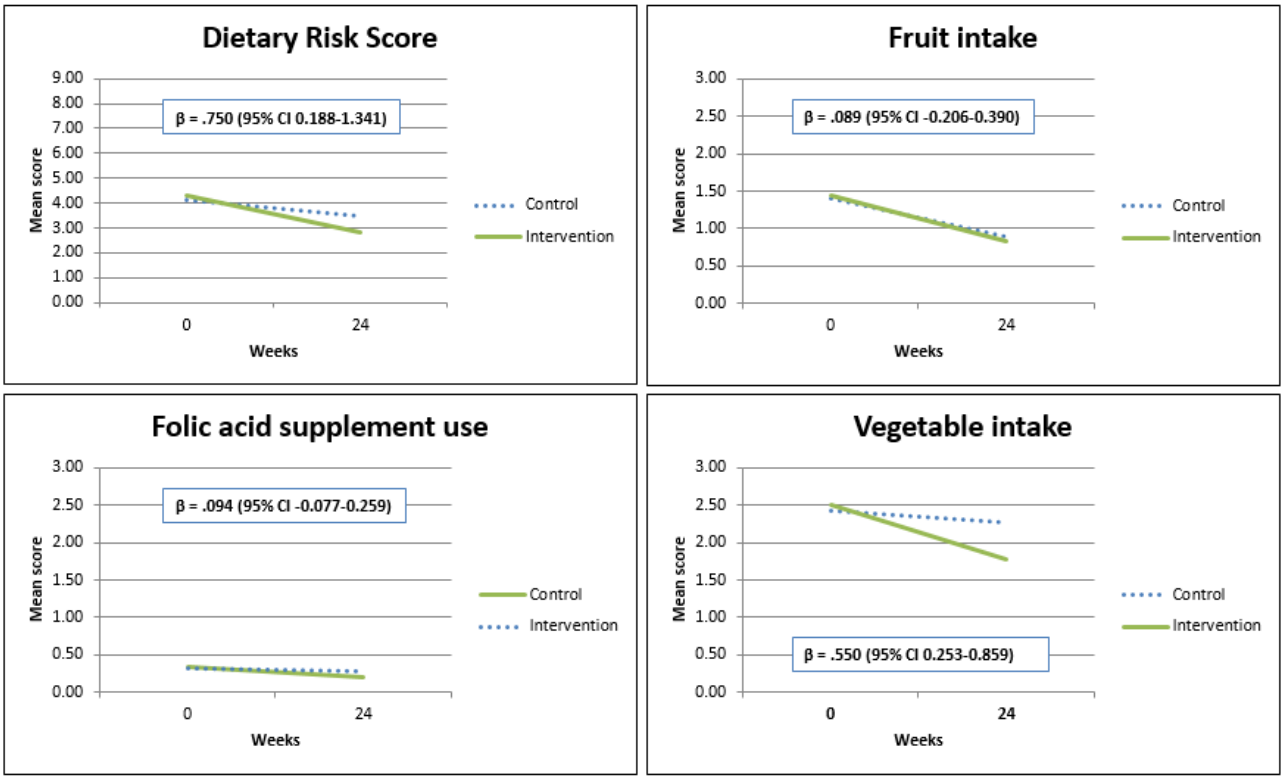
[b]BMI: body mass index.

[c]DRS: dietary risk score.

Dietary Risk Score

The outcomes of the multivariable linear regression model regarding the DRS and separate risk factors are depicted in Figure 2. Compared with participants in the control group, participants in the intervention group showed a significantly larger reduction in the DRS (β=.750; 95% CI 0.188-1.341), in particular for vegetable intake (β=.550; 95% CI 0.253-0.859). There were no significant differences between groups regarding fruit intake and folic acid supplement intake.

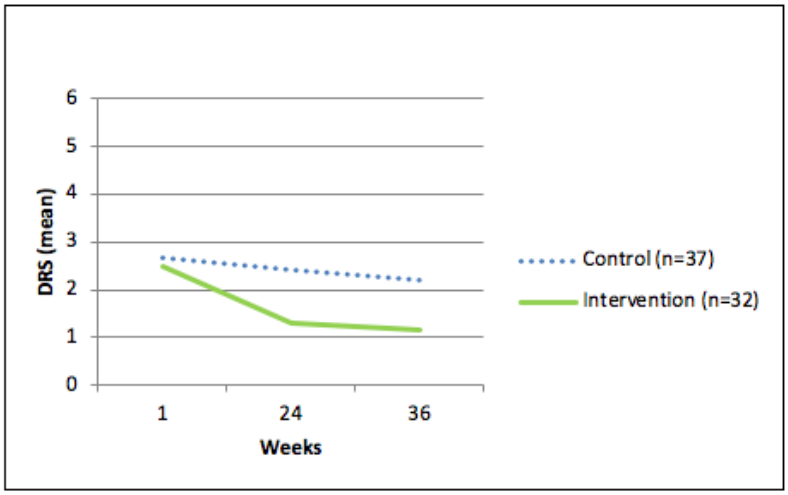
Figure 2. Dietary risk score (DRS) and separate risk factors over time in all women (with a baseline DRS>0) in the intervention and control groups. The linear regression model includes adjustment for baseline DRS and randomization.



Follow-Up

A total of 69 women (27.1%) completed the follow-up questionnaire 36 weeks after randomization. Although women did not receive any coaching during the period between 24 and 36 weeks, the mean DRS appeared to continue to decrease in both the intervention group (n=32) and control group (n=37; Figure 3). All men were lost to follow-up regarding the questionnaire at 36 weeks.

Figure 3. Mean dietary risk score (DRS) over time in all women who completed the follow-up questionnaire at 36 weeks (n=69).



Discussion

Principal Findings

The results of this randomized controlled trial indicate that compliance with the empowering lifestyle change intervention Smarter Pregnancy is high. Our findings also demonstrated that the use of this program improved the intake of vegetables (in particular) in women prior to conception and during early pregnancy. The follow-up results also suggest that this intervention had a positive effect on the continuation of healthier nutrition intake 12 weeks after stopping the intervention.

Strengths and Limitations

Major strengths of this study are the randomized controlled trial design, the fact that we used a standard and light version of the Smarter Pregnancy intervention, and the uniform collection of data in both the intervention and control group at baseline, 12,

24, and 36 weeks. By providing the control group with limited information and interaction, we encouraged participants in this group to adhere to the program, thereby ensuring high compliance and preventing dropout. The high compliance rates observed in both groups support this strategy. A further strength is that a wide range of professionals and non–health care professionals recruited women for the study, as women were approached not only during a scheduled hospital or midwifery visit, but also at children's day care centers, for example. Women who had not been personally invited could also enroll via the website we set up, thereby limiting selection bias. Additional strengths of this study are the longitudinal observations, the fact that male partners also participated, and that we collected additional information regarding lifestyle factors, educational level, geographic origin, and pregnancy status at enrollment.

In terms of weaknesses, we experienced difficulties enrolling a sufficient amount of women in the preconception period, which was the reason we expanded our inclusion criteria to include women up to 13 weeks of pregnancy. This meant that our sample size was limited, which prevented us from carrying out subgroup analyses that would have provided additional quantitative data regarding lifestyle, fertility, pregnancy course and outcome, and cost effectiveness.

We decided to reduce the total number of questionnaires in the control group from 4 to 2 because we expected that further reduction of the number of questionnaires to 1, or even 0, would have led to a higher number of resigning participants. This led to a difference in questionnaire administration, which can be considered a limitation.

A further limitation was that the program Smarter Pregnancy was only available online and in the Dutch language, thereby excluding women who have insufficient knowledge of this language or no internet access from participation in this trial. This might have excluded a high-risk population of women who might have benefitted the most from the program. An English version of this program will be released very soon.

Comparison With Other Studies

To date, there is little scientific evidence for the success of nutrition and lifestyle interventions during the preconception or periconception period. Most studies regarding preconception interventions have focused on micronutrient supplementation or weight gain, for example [36,37], or on specific subgroups and disease-related conditions, such as fertility treatment [38], polycystic ovary syndrome [39], or pre-existing/gestational diabetes [40,41]. Even fewer studies have looked at interventions on the mobile phone specifically focused on the preconception and periconception periods [40,42].

The small amount of scientific evidence regarding mHealth during the periconception period mainly showed comparable results regarding the prevalence of unhealthy nutrition and lifestyle, especially insufficient fruit and vegetable intake and smoking. These and other studies also showed and specifically addressed the relatively high dropout rates among all users [43-46]. There is a clear need for high-quality evidence that intervening in these periods in general is indeed effective, since many studies could also not demonstrate significant effectiveness [45,47,48]. A key problem underlying this lack of evidence is the lack of awareness of the importance of periconception care among both patients and health care professionals, resulting in low adherence and uptake of such care. This was described in 2002 by De Weerd et al [49]. While this barrier is widely acknowledged, and various studies have focused on how to overcome it, unfortunately barriers still remain [24,28,50,51]. It has been suggested that modern marketing campaigns such as those increasingly found online might help to overcome or at least lower this barrier [30]. Taking this into account, together with the wide uptake of mobile devices and online information [52], we believe that our approach—using a personalized intervention on the mobile phone specifically targeted at identifying and improving periconceptional risk factors—can contribute to lowering the unawareness barrier.

Conclusions and Future Perspectives

To our knowledge, Smarter Pregnancy is the first intervention on the mobile phone showing effectiveness in empowering women to improving healthy nutrition before and during early pregnancy. We therefore consider this study a good example of a successful intervention study, of which the findings support the considerable potential of using mobile phone apps. Current awareness among health care professionals of their responsibility to inform their patients about healthy nutrition is very low [53]. However, we assume that the increasing amount of evidence for the importance of nutrition in the periconception period will make health care professionals particularly more aware and make them more likely to recommend evidence-based interventions to their patients. This will contribute to an increase in the general awareness of the importance of the periconception period. As a result, we hope that periconception care will become more easily and more widely accessible, thereby improving reproductive and pregnancy outcomes in both fertile and subfertile couples.

Acknowledgments

This research was funded by the Department of Obstetrics and Gynecology, Erasmus Medical Center, University Medical Centre, Rotterdam, the Netherlands; a grant from the Netherlands Organisation for Health Research and Development (ZonMW) Health Care Efficiency Research program; and the Erasmus Medical Center Mrace Health Care Efficiency Research program. We would like to thank all women and men who participated. We also would like to thank all health care professionals involved in the recruitment of participants at the midwifery practices, hospitals, children's day care centers (KindeRdam), and child health centers in the Rotterdam region. We thank Sally Hill of Scientific Texts for critical review of the manuscript. We would also like to thank Peercode BV, Lukkien, and OGD for their technical support and the AS Watson Group for supplying incentives and vouchers.

We also thank Prof Hans Severens and Prof Hein Raat for their help in applying for funding. The analyzed dataset of this trial is available to all involved researchers and collaborators and can be provided by the corresponding author on reasonable request. To gain data access, requestors will need to sign a data access agreement.

Authors' Contributions

MRVD contributed to the design of the study, data acquisition, and writing of the manuscript. MPHK contributed to data analysis and writing of the manuscript. ECO contributed to data acquisition. SPW wrote the statistical analysis plan. EAPS was responsible for preconceptional care, facilitated the recruitment of participants, and critically reviewed the final version of the manuscript. RPMST initiated and developed the Smarter Pregnancy program and was responsible for all aspects of the study as well as contributing to all versions of the manuscript. All authors critically revised the manuscript for intellectual content and read and approved the final version.

Conflicts of Interest

Since 2016, RPMST has been Chief Executive Officer of eHealth Care Solutions. The other authors declare no competing interests.

Multimedia Appendix 1
Baseline characteristics of all men in the intervention and control groups.
[DOCX File , 15 KB - jmir_v22i5e15773_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH 2010 checklist.
[PDF File (Adobe PDF File), 53 KB - jmir_v22i5e15773_app2.pdf]

References

1. Haslam DW, James WPT. Obesity. Lancet 2005 Oct 01;366(9492):1197-1209. [doi: 10.1016/S0140-6736(05)67483-1] [Medline: 16198769]
2. Hanson MA, Gluckman PD. Early developmental conditioning of later health and disease: physiology or pathophysiology? Physiol Rev 2014 Oct;94(4):1027-1076 [FREE Full text] [doi: 10.1152/physrev.00029.2013] [Medline: 25287859]
3. World Health Organization. Global status report on noncommunicable diseases 2010 URL: http://whqlibdoc.who.int/publications/2011/9789240686458_eng.pdf?ua=1 [accessed 2010-02-10]
4. Stephenson J, Heslehurst N, Hall J, Schoenaker DAJM, Hutchinson J, Cade JE, et al. Before the beginning: nutrition and lifestyle in the preconception period and its importance for future health. Lancet 2018 May 05;391(10132):1830-1841. [doi: 10.1016/S0140-6736(18)30311-8] [Medline: 29673873]
5. Hanson M, Gluckman P. Developmental origins of noncommunicable disease: population and public health implications. Am J Clin Nutr 2011 Dec;94(6 Suppl):1754S-1758S. [doi: 10.3945/ajcn.110.001206] [Medline: 21525196]
6. World Health Organization. Action plan for the global strategy for the prevention and control of noncommunicable diseases, 2008-2013 URL: http://whqlibdoc.who.int/publications/2009/9789241597418_eng.pdf?ua=1 [accessed 2020-02-10]
7. Homan GF, Davies M, Norman R. The impact of lifestyle factors on reproductive performance in the general population and those undergoing infertility treatment: a review. Hum Reprod Update 2007;13(3):209-223. [doi: 10.1093/humupd/dml056] [Medline: 17208948]
8. Sinclair KD, Watkins AJ. Parental diet, pregnancy outcomes and offspring health: metabolic determinants in developing oocytes and embryos. Reprod Fertil Dev 2013;26(1):99-114. [doi: 10.1071/RD13290] [Medline: 24305182]
9. Sharma R, Biedenharn KR, Fedor JM, Agarwal A. Lifestyle factors and reproductive health: taking control of your fertility. Reprod Biol Endocrinol 2013 Jun 16;11:66 [FREE Full text] [doi: 10.1186/1477-7827-11-66] [Medline: 23870423]
10. Hammiche F, Laven JSE, van Mil N, de Cock M, de Vries JH, Lindemans J, et al. Tailored preconceptional dietary and lifestyle counselling in a tertiary outpatient clinic in The Netherlands. Hum Reprod 2011 Sep;26(9):2432-2441 [FREE Full text] [doi: 10.1093/humrep/der225] [Medline: 21752799]
11. Anderson K, Nisenblat V, Norman R. Lifestyle factors in people seeking infertility treatment: a review. Aust N Z J Obstet Gynaecol 2010 Feb;50(1):8-20. [doi: 10.1111/j.1479-828X.2009.01119.x] [Medline: 20218991]
12. Van Dijk MR, Huijgen NA, Willemsen SP, Laven JS, Steegers EA, Steegers-Theunissen RP. Impact of an mHealth platform for pregnancy on nutrition and lifestyle of the reproductive population: a survey. JMIR Mhealth Uhealth 2016;4(2):e53 [FREE Full text] [doi: 10.2196/mhealth.5197] [Medline: 27234926]
13. van der Pal-de Bruin KM, Le Cessie S, Elsinga J, de Jong-Potjer LC, van Haeringen A, Neven AK, et al. Pre-conception counselling in primary care: prevalence of risk factors among couples contemplating pregnancy. Paediatr Perinat Epidemiol 2008 May;22(3):280-287. [doi: 10.1111/j.1365-3016.2008.00930.x] [Medline: 18426523]
14. McPherson NO, Fullston T, Aitken RJ, Lane M. Paternal obesity, interventions, and mechanistic pathways to impaired health in offspring. Ann Nutr Metab 2014;64(3-4):231-238. [doi: 10.1159/000365026] [Medline: 25300265]

15. Augood C, Duckitt K, Templeton AA. Smoking and female infertility: a systematic review and meta-analysis. Hum Reprod 1998 Jun;13(6):1532-1539. [doi: 10.1093/humrep/13.6.1532] [Medline: 9688387]
16. Donnelly GP, McClure N, Kennedy MS, Lewis SE. Direct effect of alcohol on the motility and morphology of human spermatozoa. Andrologia 1999 Jan;31(1):43-47. [Medline: 9949888]
17. Stuppia L, Franzago M, Ballerini P, Gatta V, Antonucci I. Epigenetics and male reproduction: the consequences of paternal lifestyle on fertility, embryo development, and children lifetime health. Clin Epigenetics 2015;7:120 [FREE Full text] [doi: 10.1186/s13148-015-0155-4] [Medline: 26566402]
18. Gezondheidsraad: Health Council of the Netherlands. 2007 Sep 20. Preconception care: a good beginning URL: https://tinyurl.com/vpzwf9s [accessed 2020-02-10]
19. Steegers-Theunissen RPM, Twigt J, Pestinger V, Sinclair KD. The periconceptional period, reproduction and long-term health of offspring: the importance of one-carbon metabolism. Hum Reprod Update 2013;19(6):640-655 [FREE Full text] [doi: 10.1093/humupd/dmt041] [Medline: 23959022]
20. Jaddoe VWV, de Jonge LL, Hofman A, Franco OH, Steegers EAP, Gaillard R. First trimester fetal growth restriction and cardiovascular risk factors in school age children: population based cohort study. BMJ 2014 Jan 23;348:g14 [FREE Full text] [doi: 10.1136/bmj.g14] [Medline: 24458585]
21. Temel S, van Voorst SF, Jack BW, Denktaş S, Steegers EAP. Evidence-based preconceptional lifestyle interventions. Epidemiol Rev 2014;36:19-30. [doi: 10.1093/epirev/mxt003] [Medline: 23985430]
22. Stephenson J, Patel D, Barrett G, Howden B, Copas A, Ojukwu O, et al. How do women prepare for pregnancy? Preconception experiences of women attending antenatal services and views of health professionals. PLoS One 2014;9(7):e103085 [FREE Full text] [doi: 10.1371/journal.pone.0103085] [Medline: 25058333]
23. Temel S, van Voorst SF, de Jong-Potjer LC, Waelput AJM, Cornel MC, de Weerd SR, et al. The Dutch national summit on preconception care: a summary of definitions, evidence and recommendations. J Community Genet 2015 Jan;6(1):107-115 [FREE Full text] [doi: 10.1007/s12687-014-0204-2] [Medline: 25394755]
24. Poels M, Koster MPH, Boeije HR, Franx A, van Stel HF. Why do women not use preconception care? A systematic review on barriers and facilitators. Obstet Gynecol Surv 2016 Oct;71(10):603-612. [doi: 10.1097/OGX.0000000000000360] [Medline: 27770130]
25. M'hamdi HI, van Voorst SF, Pinxten W, Hilhorst MT, Steegers EAP. Barriers in the uptake and delivery of preconception care: exploring the views of care providers. Matern Child Health J 2017 Jan;21(1):21-28 [FREE Full text] [doi: 10.1007/s10995-016-2089-7] [Medline: 27423236]
26. Bortolus R, Oprandi NC, Rech Morassutti F, Marchetto L, Filippini F, Agricola E, et al. Why women do not ask for information on preconception health? A qualitative study. BMC Pregnancy Childbirth 2017 Jan 05;17(1):5 [FREE Full text] [doi: 10.1186/s12884-016-1198-z] [Medline: 28056865]
27. Free C, Phillips G, Felix L, Galli L, Patel V, Edwards P. The effectiveness of M-health technologies for improving health and health services: a systematic review protocol. BMC Res Notes 2010;3:250 [FREE Full text] [doi: 10.1186/1756-0500-3-250] [Medline: 20925916]
28. Van Dijk MR, Koster MP, Rosman AN, Steegers-Theunissen RP. Opportunities of mHealth in preconception care: preferences and experiences of patients and health care providers and other involved professionals. JMIR Mhealth Uhealth 2017 Aug 17;5(8):e123 [FREE Full text] [doi: 10.2196/mhealth.7834] [Medline: 28818820]
29. Krishna S, Boren SA, Balas EA. Healthcare via cell phones: a systematic review. Telemed J E Health 2009 Apr;15(3):231-240. [doi: 10.1089/tmj.2008.0099] [Medline: 19382860]
30. Barker M, Dombrowski SU, Colbourn T, Fall CHD, Kriznik NM, Lawrence WT, et al. Intervention strategies to improve nutrition and health behaviours before conception. Lancet 2018 May 05;391(10132):1853-1864 [FREE Full text] [doi: 10.1016/S0140-6736(18)30313-1] [Medline: 29673875]
31. van Dijk MR, Koster MPH, Willemsen SP, Huijgen NA, Laven JSE, Steegers-Theunissen RPM. Healthy preconception nutrition and lifestyle using personalized mobile health coaching is associated with enhanced pregnancy chance. Reprod Biomed Online 2017 Oct;35(4):453-460. [doi: 10.1016/j.rbmo.2017.06.014] [Medline: 28688924]
32. Quinn CC, Shardell MD, Terrin ML, Barr EA, Ballew SH, Gruber-Baldini AL. Cluster-randomized trial of a mobile phone personalized behavioral intervention for blood glucose control. Diabetes Care 2011 Sep;34(9):1934-1942 [FREE Full text] [doi: 10.2337/dc11-0366] [Medline: 21788632]
33. Twigt JM, Bolhuis MEC, Steegers EAP, Hammiche F, van Inzen WG, Laven JSE, et al. The preconception diet is associated with the chance of ongoing pregnancy in women undergoing IVF/ICSI treatment. Hum Reprod 2012 Aug;27(8):2526-2531 [FREE Full text] [doi: 10.1093/humrep/des157] [Medline: 22593431]
34. van Dijk MR, Oostingh EC, Koster MPH, Willemsen SP, Laven JSE, Steegers-Theunissen RPM. The use of the mHealth program Smarter Pregnancy in preconception care: rationale, study design and data collection of a randomized controlled trial. BMC Pregnancy Childbirth 2017 Dec 26;17(1):46 [FREE Full text] [doi: 10.1186/s12884-017-1228-5] [Medline: 28125970]
35. Efron B, Tibshirani R. An Introduction to the Bootstrap. London: Chapman & Hall; 1993.
36. Hillemeier MM, Downs DS, Feinberg ME, Weisman CS, Chuang CH, Parrott R, et al. Improving women's preconceptional health: findings from a randomized trial of the Strong Healthy Women intervention in the Central Pennsylvania women's

health study. Womens Health Issues 2008;18(6 Suppl):S87-S96 [FREE Full text] [doi: 10.1016/j.whi.2008.07.008] [Medline: 19059553]
37. Gaskins AJ, Chiu Y, Williams PL, Ford JB, Toth TL, Hauser R, EARTH Study Team. Association between serum folate and vitamin B-12 and outcomes of assisted reproductive technologies. Am J Clin Nutr 2015 Oct;102(4):943-950. [doi: 10.3945/ajcn.115.112185] [Medline: 26354529]
38. Kermack AJ, Calder PC, Houghton FD, Godfrey KM, Macklon NS. A randomised controlled trial of a preconceptional dietary intervention in women undergoing IVF treatment (PREPARE trial). BMC Womens Health 2014 Nov 18;14:130 [FREE Full text] [doi: 10.1186/1472-6874-14-130] [Medline: 25407227]
39. Legro RS, Dodson WC, Kris-Etherton PM, Kunselman AR, Stetter CM, Williams NI, et al. Randomized controlled trial of preconception interventions in infertile women with polycystic ovary syndrome. J Clin Endocrinol Metab 2015 Nov;100(11):4048-4058 [FREE Full text] [doi: 10.1210/jc.2015-2778] [Medline: 26401593]
40. Charron-Prochownik D, Ferons-Hannan M, Sereika S, Becker D. Randomized efficacy trial of early preconception counseling for diabetic teens (READY-girls). Diabetes Care 2008 Jul;31(7):1327-1330 [FREE Full text] [doi: 10.2337/dc07-1266] [Medline: 18411239]
41. Koivusalo SB, Rönö K, Klemetti MM, Roine RP, Lindström J, Erkkola M, et al. Gestational diabetes mellitus can be prevented by lifestyle intervention: the Finnish Gestational Diabetes Prevention Study (RADIEL): a randomized controlled trial. Diabetes Care 2016 Jan;39(1):24-30. [doi: 10.2337/dc15-0511] [Medline: 26223239]
42. Nwolise CH, Carey N, Shawe J. Preconception care education for women with diabetes: a systematic review of conventional and digital health interventions. J Med Internet Res 2016 Nov 08;18(11):e291 [FREE Full text] [doi: 10.2196/jmir.5615] [Medline: 27826131]
43. van Zutphen M, Milder IE, Bemelmans WJ. Usage of an online healthy lifestyle program by pregnant women attending midwifery practices in Amsterdam. Prev Med 2008 Jun;46(6):552-557. [doi: 10.1016/j.ypmed.2008.01.003] [Medline: 18302971]
44. Bot M, Milder IEJ, Bemelmans WJE. Nationwide implementation of Hello World: a Dutch email-based health promotion program for pregnant women. J Med Internet Res 2009;11(3):e24 [FREE Full text] [doi: 10.2196/jmir.1183] [Medline: 19674957]
45. Choi J, Lee JH, Vittinghoff E, Fukuoka Y. mHealth physical activity intervention: a randomized pilot study in physically inactive pregnant women. Matern Child Health J 2016 May;20(5):1091-1101. [doi: 10.1007/s10995-015-1895-7] [Medline: 26649879]
46. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
47. Tieu J, Middleton P, Crowther CA, Shepherd E. Preconception care for diabetic women for improving maternal and infant health. Cochrane Database Syst Rev 2017 Aug 11;8:CD007776 [FREE Full text] [doi: 10.1002/14651858.CD007776.pub3] [Medline: 28799164]
48. Brown HK, Mueller M, Edwards S, Mill C, Enders J, Graves L, et al. Preconception health interventions delivered in public health and community settings: a systematic review. Can J Public Health 2017 Nov 09;108(4):e388-e397. [doi: 10.17269/cjph.108.6029] [Medline: 29120310]
49. de Weerd S, Steegers EAP. The past and present practices and continuing controversies of preconception care. Community Genet 2002;5(1):50-60. [doi: 10.1159/000064631] [Medline: 14960900]
50. Elsinga J, de Jong-Potjer LC, van der Pal-de Bruin KM, Le Cessie S, Assendelft WJJ, Buitendijk SE. The effect of preconception counselling on lifestyle and other behaviour before and during pregnancy. Womens Health Issues 2008 Dec;18(6 Suppl):S117-S125. [doi: 10.1016/j.whi.2008.09.003] [Medline: 19059545]
51. Poels M, Koster MPH, Franx A, van Stel HF. Parental perspectives on the awareness and delivery of preconception care. BMC Pregnancy Childbirth 2017 Sep 26;17(1):324 [FREE Full text] [doi: 10.1186/s12884-017-1531-1] [Medline: 28950838]
52. Tan SS, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. J Med Internet Res 2017 Jan 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
53. van Voorst S, Plasschaert S, de Jong-Potjer L, Steegers E, Denktaş S. Current practice of preconception care by primary caregivers in the Netherlands. Eur J Contracept Reprod Health Care 2016 Jun;21(3):251-258. [doi: 10.3109/13625187.2016.1154524] [Medline: 27003266]

Abbreviations

BMI: body mass index
DRS: dietary risk score
eHealth: electronic health
mHealth: mobile health
NCD: noncommunicable disease
ZonMW: Netherlands Organization for Health Research and Development

Edited by G Eysenbach; submitted 05.08.19; peer-reviewed by E Hewawasam, A Rosman; comments to author 24.09.19; revised version received 30.10.19; accepted 16.12.19; published 15.05.20.

Please cite as:
van Dijk MR, Koster MPH, Oostingh EC, Willemsen SP, Steegers EAP, Steegers-Theunissen RPM
A Mobile App Lifestyle Intervention to Improve Healthy Nutrition in Women Before and During Early Pregnancy: Single-Center Randomized Controlled Trial
J Med Internet Res 2020;22(5):e15773
URL: https://www.jmir.org/2020/5/e15773
doi:10.2196/15773
PMID:32412417

Original Paper

The Development of a Web-Based, Patient-Centered Intervention for Patients With Chronic Myeloid Leukemia (CMyLife): Design Thinking Development Approach

Geneviève ICG Ector[1], MSc, MD; Peter E Westerweel[2], MD, PhD; Rosella PMG Hermens[3], PhD; Karin AE Braspenning[3], MSc; Barend CM Heeren[4], MSc; Oscar MF Vinck, MSc; Jan JM de Jong[5], MSc, MD; Jeroen JWM Janssen[6], MD, PhD; Nicole MA Blijlevens[1], MD, PhD

[1]Department of Hematology, Radboud University Medical Center, Nijmegen, Netherlands

[2]Department of Hematology, Albert Schweitzer Hospital, Dordrecht, Netherlands

[3]Department of IQ Healthcare, Radboud University Medical Center, Nijmegen, Netherlands

[4]REshape Innovation Center, Radboud University Medical Center, Nijmegen, Netherlands

[5]Dutch Patient Advocacy Group 'Hematon', Utrecht, Netherlands

[6]Department of Hematology, Cancer Center Amsterdam, Vrije Universiteit Amsterdam, Amsterdam University Medical Centers, Amsterdam, Netherlands

Corresponding Author:
Geneviève ICG Ector, MSc, MD
Department of Hematology
Radboud University Medical Center
Geert Grooteplein Zuid 8
Nijmegen, 6500 Hb
Netherlands
Phone: 31 243618810
Email: genevieve.ector@radboudumc.nl

Abstract

Background: With the global rise in chronic health conditions, health care is transforming, and patient empowerment is being emphasized to improve treatment outcomes and reduce health care costs. Patient-centered innovations are needed. We focused on patients with chronic myeloid leukemia (CML), a chronic disease with a generally good long-term prognosis because of the advent of tyrosine kinase inhibitors. However, both medication adherence by patients and guideline adherence by physicians are suboptimal, unnecessarily jeopardizing treatment outcomes.

Objective: The aim of this study was to develop a patient-centered innovation for patients with CML using a design thinking methodology.

Methods: The 5 phases of design thinking (ie, empathize, define, ideate, prototype, and test) were completed, and each phase started with the patient. Stakeholders and end users were identified and interviewed, and observations in the care system were made. Using tools in human-centered design, problems were defined and various prototypes of solutions were generated. These were evaluated by patients and stakeholders and then further refined.

Results: The patients desired (1) insights into their own disease; (2) insights into the symptoms experienced, both in terms of knowledge and comprehension; and (3) improvements in the organization of care delivery. A web-based platform, CMyLife, was developed and pilot-tested. It has multiple features, all targeting parts of the bigger solution, including a website with reliable information and a forum, a guideline app, personal medical records with logs of symptoms and laboratory results (including a molecular marker and linked to the guideline app), tailored feedback based on the patients' symptoms and/or results, screen-to-screen consulting, delivery of medication, and the collection of blood samples at home.

Conclusions: The multifeatured innovation, CMyLife, was developed in a multidisciplinary way and with active patient participation. The aim of developing CMyLife was to give patients the tools to monitor their results, interpret these results, and act on them. With this tool, they are provided with the know-how to consider their results in relation to their personal care process. Whether CMyLife achieves its goal and the evaluation of the added value will be the focus of future studies. CML could become the first malignancy for which patients are able to monitor and manage their disease by themselves.

(J Med Internet Res 2020;22(5):e15895) doi:10.2196/15895

KEYWORDS

eHealth; chronic myeloid leukemia; patient participation; mobile apps

Introduction

Background

In 2002, the World Health Organization reported about expecting dramatic changes in global health because of the rising prevalence of chronic health conditions, underscoring the necessity for innovations in health care to effectively manage long-term health problems [1], which also requires a shift in the traditional roles of physicians and patients. The growing involvement of patients in their own care process and self-management needs to be stimulated [2] as this could improve the health outcomes of treatment and reduce health care costs [3]. Ideally, care for chronic illnesses should be patient-centered, resulting in an increased understanding and awareness of health, treatment options, symptoms, and behaviors [4,5]. This patient-centeredness should be of major focus when facilitating innovation in health care. In other sectors, involving the end user of a product in its development is a common practice, and different kinds of human-centered approaches are available [6]. One of these approaches is *design thinking*. Although a universal definition is missing, the solution-focused methodology is best described as "a discipline that uses the designer's sensibility and methods to match people's needs with what is technologically feasible and what a viable business strategy can convert into customer value and market opportunity" [7]. By engaging the end user from the beginning of the process, solutions are developed bottom-up instead of top-down [8]. Nowadays, design thinking is also gaining ground in the field of health care and health care education [9-13]. Design thinking and user-centered design are quite similar and start from the same standpoint—the user. The steps of both processes are quite comparable. However, design thinking emphasizes the ideation phase and focuses on finding solutions and has a specific set of tools.

Objective

This paper aimed to describe the development and pilot testing of a patient-centered innovation that can change the current model of health care. As starting with small and rapid prototyping is key, we started with changing care for a specific group of patients (ie, patients with chronic myeloid leukemia [CML]). This malignant disease results from a reciprocal translocation between chromosomes 9 and 22, generating an abnormal Philadelphia chromosome. The resulting *BCR-ABL1* fusion gene encodes for a constitutively active tyrosine kinase, which is pivotal in the pathogenesis of CML and can be inhibited by tyrosine kinase inhibitors (TKIs). With the advent of TKIs, CML became a chronic disease rather than an often fatal one [14,15]. As with many chronic diseases, therapy compliance and guideline adherence are challenging. The paramount step in the monitoring of patients with CML is the quantification of *BCR-ABL1* transcripts with a polymerase chain reaction on peripheral blood, which represents disease activity [16]. The recommendations for the monitoring of patients with CML are defined in international guidelines [16,17]. However, guideline adherence by physicians is often suboptimal [18-20], which is associated with a higher risk of progression to an advanced phase of the disease and death [21]. Another challenge in CML care comprises suboptimal adherence to TKIs by patients, which in part relates to the adverse events (AEs) caused by TKIs [18,21-24]. Common AEs are muscle aches and cramps, fatigue, gastrointestinal symptoms, and fluid retention, among others [25,26]. AEs are the main cause of intentional nonadherence. The adequate management of AEs is a necessity, as an adherence rate of ≤90% is clearly related to a lower probability of reaching important treatment goals [22]. Furthermore, physicians tend to underestimate the severity of AEs experienced by patients and overestimate the patients' overall health status [27].

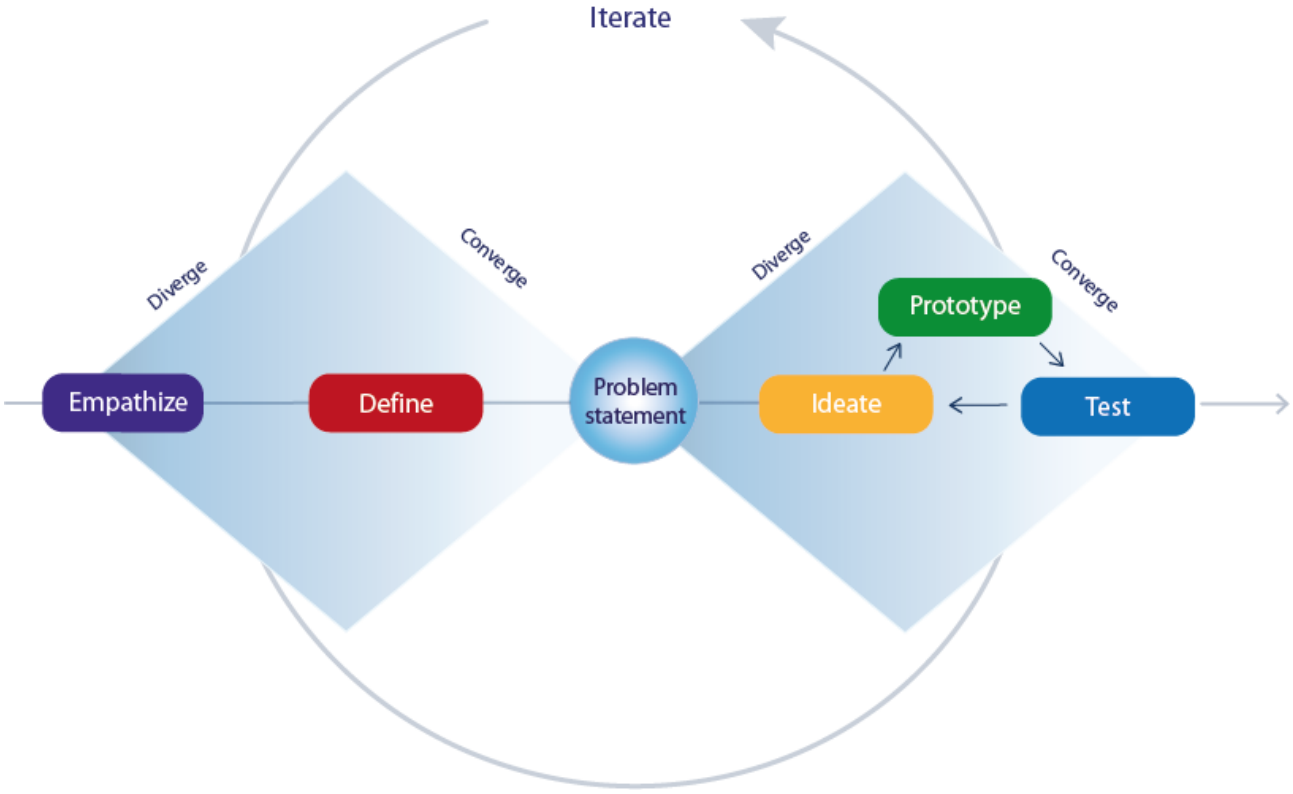
By transforming CML care into a patient-centered care model, the self-management of monitoring and intake of TKI can be stimulated, resulting in improved health outcomes and the involvement of patients in their own care process. Hence, CML is pre-eminently suited for developing an innovative patient-centered care model.

Methods

Overview of the Study Design

The design thinking methodology [28] was used for the development of the innovative patient-centered care model. It comprises the following 5 phases: empathize, define, ideate, prototype, and test. We started with the key element of every phase, ie, the end user, and in our case, the patient. Pilot evaluations of the developed innovation were performed by focus group interviews with patients. Design thinking is a dynamic process where phases can run parallel or in iterative circles based on new insights or feedback. Going back is possible at any moment during the process. Generally, it can be visualized in the shape of 2 diamonds: first diverge in empathizing, then converge and define a problem, then again diverge and converge, generating ideas for a solution and prototyping and testing the right solution, respectively. Figure 1 represents this process and is adapted from an illustration developed by Jasper Liu [29].

Figure 1. The double diamond model depicting the essentials of the design thinking process.



Setting

The project has been established by a collaboration among hematologists at 3 hospitals in the Netherlands, the Dutch hematological patient advocacy group, *Hematon*, and the Reshape and Innovation Center of Radboudumc. A multidisciplinary project team was formed to coordinate the process, comprising patients, health care professionals, designers, developers, and a communication specialist, thereby representing both the participating and the developing parties. The study was approved by the institutional review board of the Radboudumc.

Phase 1: Empathize

The understanding of the end users and their needs, desires, and challenges in everyday life is crucial and should be explored. To do so, we conducted in-depth interviews with patients and hematologists until data saturation was reached. The interviews were conducted by the user-centered designers of the project team. The patients were asked to participate during a national meeting of the patient association *Hematon* or were approached by their hematologists. The hematologists were approached via professional associations. Furthermore, field observations of different parts of the patient's journey in the health care system were made. These observations took place in pharmacies, in waiting rooms, during blood sample collection, and at the outpatient clinics during consultations with the hematologists. A checklist was used to evaluate a certain set of items (eg, providing information on prescription medication, waiting time). Communication science students conducted the observations and were supervised by the project team. All information gathered in this phase was clustered into themes to identify connections. Furthermore, a stakeholder analysis was made, with the aid of stakeholder mapping, to identify all stakeholders.

Phase 2: Define

During this analyzing phase, all information gathered in the *empathizing phase* was translated into a problem statement: an expression of the problem we were aiming to address. The statement had to be human-centered, broad enough for creative freedom and narrow enough to make it manageable [30]. Therefore, tools such as affinity clustering, the 5 whys, frameworks, and creating a point of view were used [31-33]. The statement was defined by the project team in an iterative process until consensus was achieved.

Phase 3: Ideate

Here, the goal was to generate the broadest range of ideas for design as possible, preferably from different stakeholders' perspectives. We started ideating with a selection of the project team comprising at least a designer, patient, developer, and health care professional. Subsequently, the ideas were shared and modified by the entire project team. By asking "how might we" questions, the information gathered in the previous phases was converted into requirements of the innovation. It was key to strive for as many ideas as possible to solve parts of the problems and not just for one idea to solve the whole problem. In this manner innovation concepts could be formed and tested on a small scale.

Phase 4: Prototype

We started with rapid prototyping by developing many low-resolution (quick to make and without much expenses) prototypes, eg, screenshots of possible apps and websites. These were tested early in the process and immediately refined after

feedback from the end users and stakeholders. Feedback was collected during the test sessions with small groups of end users. In addition, meetings were held with the project team, complemented by the representatives of different stakeholders. If additional information was found to be missing or new ideas came up, the previous phases were repeated.

Phase 5: Test

This phase ran partially parallel with the prototype phase where every feature from the mock-up to concept product was tested iteratively with the end users. Eventually, a pilot test was run on a small scale at the Radboudumc (the center where the project team was situated). A stepwise workflow was developed, containing the tasks and requirements that had to be fulfilled at each step and the person who had to fulfill these in the first 12 months. To create awareness of the innovation, a communication plan was developed for both patients and health care professionals. In the prototype phase, the prototypes were tested in small groups of mostly patients (4-6 patients). During the pilot test, focus groups were conducted with the help of a structured interview guide (Multimedia Appendices 1 and 2). They were conducted with groups of patients and hematologists. Participants were approached via the website of the innovation, *Hematon*, and professional associations. The first focus group comprised 4 patients and 1 patient's caregiver (3 males and 2 females, aged 54-72 years). In the subsequent focus group, 7 patients participated. They were interviewed about the introduction to and awareness of the innovation, practical issues in utilizing it, layout, content, and the added value in one's own life. Another focus group was conducted with 13 hematologists, following a CML-focused national working group. The topics were similar to those in the patients' focus groups (see Multimedia Appendix 2 for the interview guide). Quantitative data on the usage of the innovation's website and different mobile features were established at 12 and 16 months using Google Analytics and data from the third parties involved, respectively.

Results

Phase 1: Empathize

Overall, 8 patients (data saturation was reached) were interviewed to understand their experiences and explore their needs and desires. The mean age of these patients was 53 years (range 28-75), and 75% (6/8) were male. Each interview lasted between 60 and 90 min. The patients' strongest wish was to be cured of the disease; however, 3 other wishes were mentioned almost unanimously. First, insights into their disease, with more knowledge and comprehension, were desired (for instance, "Am I doing well or not?" and "Where am I in my disease course?"). Second, they wished for better support in understanding and coping with the symptoms they experienced (eg, "I'm having a hard time summarizing all the symptoms of the past 3 months before consultation" and "How can I explain [to] my family and friends what the impact is on my daily life?"). Finally, they wished for improvements in the organization of care delivery. The journey to the hospital and different appointments (ie, hematologist, nurse specialist, blood sample collection, pharmacy) at different times and multiple locations were considered to be time-consuming, apart from the time spent in a waiting room.

A total of 14 hematologists (9 working at a university hospital and 5 at a nonuniversity hospital) were interviewed, and they mentioned the following desires: to empower the patient, better insight into the patient-reported experiences and outcomes, improvement of guideline adherence, and providing care only when medically needed or when desired by the patient.

In addition, 34 observations (data saturation was reached) in the outpatient clinic, during blood testing, or in the pharmacy were made to fully understand the patients' journey. The observations were not connected to the patients interviewed and took place during different parts of the day. Lengthy waiting times were the biggest issue in almost all observations as well as receiving too little information regarding the prescription medication at the pharmacy. Altogether, the role of the patients was too small in their own care process, and adequate tools to take the lead were lacking.

The following stakeholders were identified: the patients as end users; health care professionals such as hematologists, general practitioners, and, to a lesser extent, pharmacists; molecular biologists; and health insurers.

Phase 2: Define

The objective of the innovation was defined as "to empower patients and facilitate them to take the lead in their own care process." As CML cannot be cured, we could not meet the biggest desire of curing the disease. However, it has recently become clear that it is feasible for patients to achieve a durable remission of the disease after discontinuation of their treatment [34,35]. This is referred to as *treatment-free remission* (TFR) and is considered to represent an *operational cure*. TFR is having a stable, deep molecular response without any ongoing TKI treatment. Patients in chronic phase CML who have maintained a stable and deep molecular response (for at least two years) are considered for TKI discontinuation. In an international consensus statement of patient representatives and doctors, TFR was adopted as a desirable goal of CML therapy in the context of adequate medical and psychological support [36]. It was established that optimizing guideline adherence and therapy compliance are prerequisites for achieving this goal.

Phase 3: Ideate

On the basis of the Define phase, the innovation should provide reliable information and instruments for the patients to gain insights into their own disease course and educate them in the subsequent steps. In addition, the innovation had to support patients when experiencing symptoms, whether disease-related or TKI-related. To reduce logistic struggles such as waiting times and to achieve availability for every patient with CML, the multidisciplinary project team decided that the innovation had to transfer part of the current care from the hospital to the personal environment of the patient.

Owing to the variety in problems, a variety of solutions were proposed, for instance, electronic health (eHealth) technologies to enhance knowledge and comprehension of the disease, provide a way to report and log symptoms, facilitate contact

among patients for both social support and sharing insights, improve the organization of care delivery, and eliminate unnecessary steps, and, in addition, a nurse specialist in both CML and electronic nursing, who can empower and educate patients with the aid of eHealth technologies. The multidisciplinary project team came up with a web-based innovation covering these different solutions to achieve the main goal of empowering the patient and making CML care more patient-centered by improving the quality of life, medication adherence, guideline adherence, treatment outcomes, and reducing organizational burdens of the care process.

Phase 4: Prototype

Low-profile prototypes varying in content and layout were made of various parts of the coordinating solution (eg, different graphs depicting *BCR-ABL1* values to select the most insightful one; Figure 2). Eventually, the iterative testing of the prototypes with the patients' focus groups resulted in the concept of a web-based platform called CMyLife, literally meaning *see my life* and referring to CML. The platform was extended by adding different features, with the patients deciding which features to use in their own care process. It comprised a public website containing reliable information on the disease; treatment (eg, AE profiles per TKI and TFR); other medical issues; the impact on social aspects, eg, sports and financial issues (Figure 3); and a patient-tailored part with additional features in a secure environment. These features have been described as follows.

Figure 2. Four examples of graphs depicting BCR-ABL1 values. Graphs B, C, and D have a colored area under the curve, in contrast to graph A. Furthermore, the percentages on the y-axis differ: graphs A, C, and D have percentages from 100 to ≤0.001, representing the number of BCR-ABL1 transcripts. Graph B has a positive approach with percentages from 9 to ≥99.999, representing the amount of healthy cells.

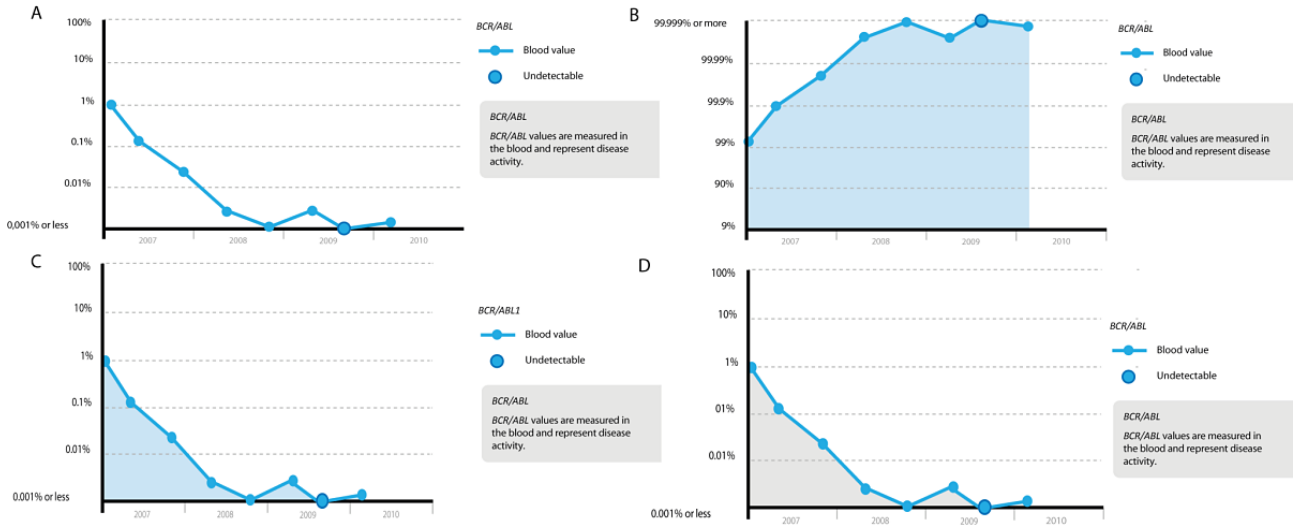
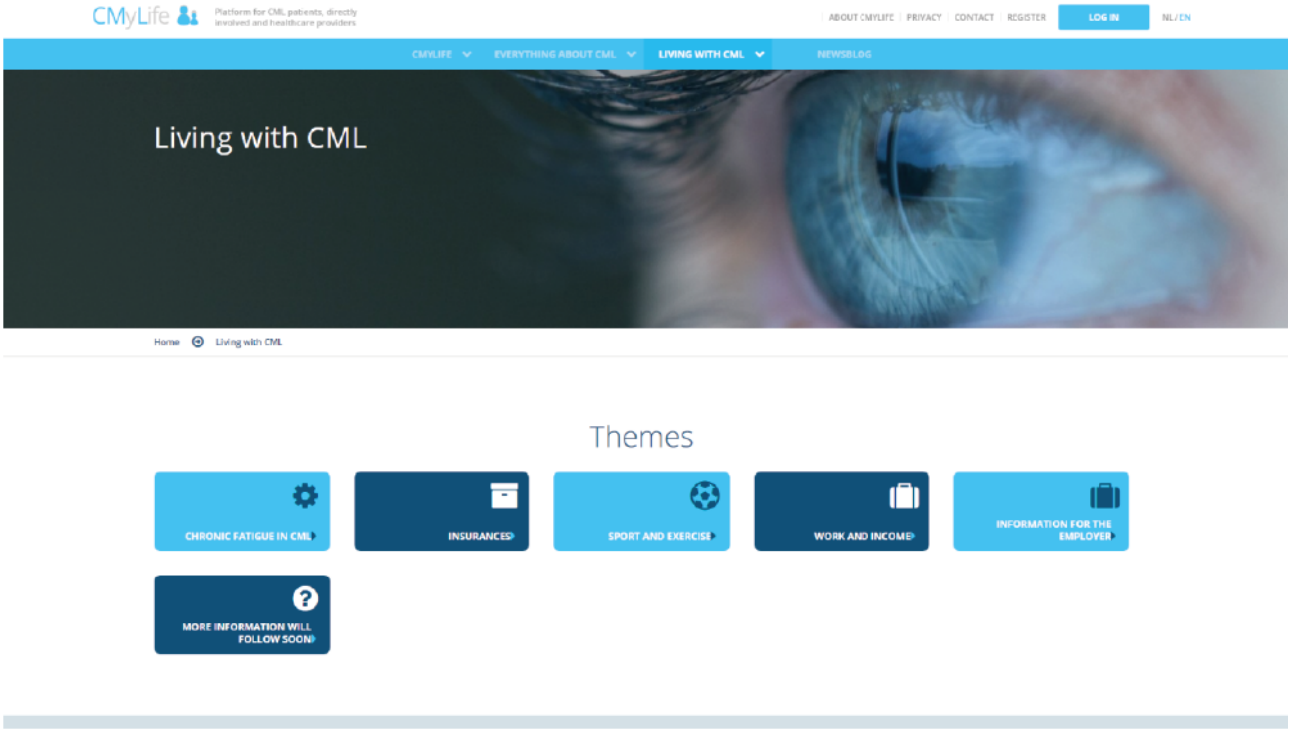


Figure 3. Screenshot of the CMyLife website.



Forum

A forum was developed in a secured part of the platform, where patients could meet other patients and ask questions. It was monitored by 2 patient-experts and the community manager of the project team who can refer the questions to a hematologist, in case of medical questions.

Patients Know Best

Another feature is the Patients Know Best portal [37], where patients are in control of their personal medical record. They can grant access to the medical records to their health care professionals (including professionals other than hematologists) or choose to keep them private. For patients in the pilot hospital, a connection was made between the hospital's eHealth record and the Patients Know Best portal. In this way, laboratory results are uploaded to the Patients Know Best automatically.

MedApp

We reached out to the developers of the mobile phone app, MedApp, aimed at helping patients managing their medications [38]. This resulted in a *CMyLife module* that is linked to a patient's Patients Know Best portal in the app. In this module, patients can rate 10 frequent TKI-related symptoms daily on a 4-point Likert scale, resulting in a weekly or monthly overview. This can be used as a visual representation of the impact of CML on their daily life and can be shared with family and friends. If the patients share their Patients Know Best record with the hematologist, this can result in a more tailored consultation. Preparations are made to develop a feedback loop signaling the nurse specialist to contact the patient in case of grade 3 or 4 AEs.

Disease Activity

To provide insight into an individual's disease course, graphs depicting the molecular marker measured to monitor disease activity (*BCR-ABL1* transcript values) are uploaded in the Patients Know Best record, in collaboration with the local molecular laboratory (Figure 4).

Figure 4. Screenshot of the graph depicting BCR-ABL1 transcript values in a patient's own Patients Know Best record in the CMyLife environment. When a dot is selected, it shows the results and its date.

Guideline Application

The above-mentioned graph (Figure 4) visualizing the molecular marker levels is linked to the guideline app. This app contains an easy to understand stepwise explanation of the Dutch CML guideline for patients with CML in chronic phase using a first-line TKI [39]. On the basis of their last molecular response and moment of testing, patients receive advise, for example:

> *The response is optimal, you should continue with your medication.*
>
> *The response is not optimal. This can be due to medication interactions, skipping medication or progression of disease. Ask your hematologist about it on your next appointment.*

In addition, reminders are sent when patients should be tested again. Expanding the functionalities to second-line treatment and CML in other phases is planned for the near future.

Reducing Hospital Visits

Furthermore, organizational issues were addressed by the following features, all aimed at minimizing hospital visits.

Blood Samples

In collaboration with the Thrombosis Service, a service monitoring anticoagulant therapy in the Netherlands, which draws blood samples at home or nearby for patients with thrombosis using oral anticoagulants, patients with CML can get their blood samples drawn at home.

Pharmacy

TKIs are delivered to the patient's home by the pharmacy. In the MedApp's CMyLife module, patients can request the pharmacy directly for a new delivery when they are almost out of TKIs. Hence, a 12-month valid prescription is supplied by the hematologist.

Screen-to-Screen Consulting

Finally, patients are offered a screen-to-screen consultation with their hematologists. Consultation rooms were provided with the right equipment, and health care professionals were trained. Face-to-face consultations with the hematologist were reduced to once a year unless additional face-to-face physical consultations at the hospital were considered desirable by the patient. Between these visits, patients were supported through screen-to-screen consultations and the other features described above.

Security

All data were secured, and data protection conformed to the national and European privacy regulations (General Data Protection Regulation). The features containing data from the hospital's electronic health record (ie, laboratory results in their personal health record, website, and app) were secured with a password-protected log-in and subsequent token via an SMS text message for authentication. This is in accordance with the Dutch security guideline NEN7510. The forum is password protected.

Before the connection can be made between the hospital's systems and CMyLife, a patient has to register for a personal health record. Subsequently, their identity is checked based on

name, date of birth, postal code, and citizen service number. The personal medical record is International Organization for Standardization/International Electrotechnical Commission 27001:2013 certified.

Phase 5: Test

Focus Groups

Patients had different experiences of how and when CMyLife was introduced to them. Some preferred an introduction by the patient association *Hematon*, others by the hematologist. With regard to the practical issues in its utilization, it turned out that patients had little experience with the platform at this time ("I did not even know it existed"). A few content-specific suggestions were made, such as a small piece of information helping patients explain CML and its impact to others. Screen-to-screen consultations and having blood samples drawn at home received a positive response:

> *I really like the video-consulting, this way I am certain I don't have to waste a half day every 3 months or so.*
>
> *My consultations with my doctor never last longer than 3 minutes. "Everything looks fine, great, bye and see you next time." (...) What am I doing here. There are people who need the hematologist more than I do, especially when I'm doing well. I would prefer video-consulting, it saves time for the both of us.*
>
> *I receive a prescription from my doctor, go to the hospital pharmacy and a few days later I can pick up my TKIs, if I am lucky...And you always have to make a call on forehand to check if your order is ready. It is old-fashioned.*

Insights into individual laboratory results was desirable when combined with an easy-to-understand version of the guidelines ("We do not want our care to be dependent of which doctor is treating you, it should be the same everywhere"). The responses to registering symptoms and medication intake were variable as everyone could not imagine this to be beneficial to their management:

> *It involves extra steps and it can result in more work, though.*
>
> *I understand the advantage, a doctor can prepare the consultation.*

Hematologists were a little more reluctant regarding CMyLife. They could imagine the added value in improving patient empowerment and transferring care to outside the hospital. However, there were concerns regarding time investment, as time is a scarce resource in daily practice:

> *I think it is very important to provide care at home and for a patient to take the lead. The problem is that as a doctor, I don't want to have extra work, so it has to be very well facilitated.*
>
> *I think it can help improving adherence. Of course, patients only experience CML-related symptoms if it went wrong, if patients monitor their own results and see the numbers and decrease in log value, it could be stimulating to be really adherent to their medication.*
>
> *All our patients search the internet for information. What you want as doctors is provision of trustworthy information, I think that is an important goal of such a CMyLife intervention.*
>
> *I see potential; however, I also foresee many technical issues in connecting the CMyLife system to our hospital system.*
>
> *Through a system like this, you can diminish a lot of routine follow-up visits every 3 months, in order to create more time for other things.*

Another topic of concern was the growing development of (commercial) mobile interventions, all specified to just a specific disease and often focused on just one aspect of that disease:

> *An important barrier is the coordination with similar initiatives for other patients than CML patients. One universal solution is desirable. The strength of CMyLife is that it can be regarded as a blueprint that can be used for other diseases.*
>
> *Patients have comorbidities, some patients see a hematologists, cardiologist and other doctors, it would be preferable if CMyLife was not a stand-alone application.*

User Statistics

With regard to the user statistics of the CMyLife website, in the first year (December 2016-December 2017), CMyLife had 6334 unique visitors, with approximately 600 visitors per month for the last 6 months. Unique page views in the last 12 months were 36,300. Per session, 5.17 pages were visited on average, with the page containing the latest blogs and information on AEs having the most visits after the homepage and log-in page. A total of 338 users were registered, who were treated in 62 different hospitals. Per month, the forum was visited on an average of 150 times by 40 users. One-fifth of the visitors were new; 80% comprised returning visitors. A total of 677 users registered for the CMyLife module in MedApp, with 156 active users (ie, monthly use) in the last month.

Running and Expanding the Pilot

The publicly available website and online user forum were launched nationwide during *Hematon*'s information day on CML, on April 23, 2016. Gradually, we extended the other features for patients of the Radboudumc to test and refine CMyLife and prepare for further implementation. The workflow of the electronic consultations has been described hereafter. As the pilot was confined to patients of the Radboudumc, the additional features (ie, except the website and user forum) were separated from the public version of the tool.

Patients received a letter from their hematologist accompanied by a CMyLife information brochure. Extra time was scheduled for their next follow-up visit. This visit comprised 4 parts. First, the hematologist completed the routine follow-up and then explained the changing model of care. If desired, a 12-month valid prescription was written for TKIs to be delivered to the patient's home. Second, an introduction with the nurse specialist

took place, followed by an explanation of the functions and first half of the features of CMyLife (information source and Patients Know Best). Third, the patient visited a member of the project team who showed an explanatory video and installed and activated CMyLife and Patients Know Best (after an identification check). If necessary, instructions were provided by email and telephone. In total, 80 min were reserved for these 3 parts. Finally, the patient visited the pharmacy where the type of medication and the number of pills in possession of the patient were registered in CMyLife. After 3 months, the patient had a routine follow-up scheduled with the hematologist, who could discuss the symptoms reported in CMyLife and TKI adherence. Thereafter, the patient visited the nurse specialist who evaluated the experiences of using CMyLife and explained the second half of the features (ie, video consulting, having blood drawn at home, and scoring symptoms). Both visits lasted for 20 min each. After 6 months, the hematologist proposed the possibility of screen consulting for the next visit, followed by an explanatory visit, including instructions for video consulting with the nurse specialist.

Discussion

Principal Findings

Finally, with the aid of the design thinking methodology, the CMyLife innovation was developed, which is a web-based platform with multiple features such as a website with reliable information and a forum, a guideline app, personal medical records with logs of symptoms and laboratory results (including molecular markers connected to the guideline app), screen-to-screen consulting, and delivery of medication and collection of blood samples at home. Since the beginning of the pilot, 338 users were registered, and the website had 6334 unique visitors. However, it should be noted that CML is a rare disease, with approximately 160 new patients per year in the Netherlands.

Gaining insights into the course of one's disease was most desired by patients. Hence, an easy-to-understand version of the Dutch CML guideline connected to personal laboratory results was developed, which provides patients with the tools to take the lead in their own care process, enhancing patient empowerment and guideline adherence. In daily practice, adherence to guidelines by physicians is often suboptimal [18-20,40,41], which is associated with lower medication adherence and inferior treatment outcomes, including an increased risk of progression to an advanced phase of the disease [19-21,24]. A second feature is the provision of trustworthy information on a variety of subjects related to CML care, based on patients' feedback. Furthermore, by transferring the majority of care outside of the hospital, CML care is centralized digitally. This could be beneficial as adherence to guidelines was found to be the lowest in small centers treating just one or a few newly diagnosed patients with CML [42]. Even for patients who have achieved TFR and discontinued TKI treatment, CMyLife could be of added value. As approximately half of the patients remain in TFR after TKI discontinuation [43-45], optimal monitoring is important.

Another key feature of CMyLife is the personal health record that can be created and filled with patient-reported outcomes (PROs) such as TKI intake, health status, and AEs. In the literature, AEs are frequently reported and are even described as the main reason for skipping TKI doses intentionally [46-49]. These low adherence rates are a serious challenge in CML care [48,50,51], leading to decreased treatment responses and subsequently increased rates of resistance and progression [22,23,47,52]. Obviously, recognition of AEs by the physician is essential to relieve the symptoms or consider switching TKI to enhance adherence. However, physicians tend to rate symptom severity lower than patients and overestimate the adherence of their patients [27,53]. Patients reporting their own outcomes have improved control of symptoms, medication adherence, and patient satisfaction. Furthermore, PROs provide distinct prognostic information, exceeding clinical measurements [54-56].

Naturally, the utilization of human-centered design approaches and the development of mobile health technologies in health care are not new. There are numerous examples of mobile health technologies for patients and their caregivers for various diseases. Even interactive patient portals in breast and lung cancer care, developed with active patient participation, are described in the literature [57,58]. Laboratory result, PRO, and frequently asked question sections are common features of these portals. Other web-based interventions include a previsit communication tool for malignant lymphoma, an online community for adolescents and young adults with cancer, online communities for patients with breast cancer, and a web-based app supporting cancer survivors in self-management by monitoring PROs and acquiring feedback with a summary of personalized supportive care options [59-62]. As such, the individual web-based features of CMyLife are not unique; however, the combined features and the transfer of care from the hospital to the patient's own environment are unique. CMyLife changes the current delivery of care and centralizes it digitally.

Laboratory results, PROs, electronic consulting, and information provision can be considered as generic components of innovative apps for all types of cancer and even for nonmalignant chronic diseases. It is key to the development of such innovations to start at the beginning, ie, with the patients, and tailor the innovation as a solution to the patients' needs. CMyLife in its entirety incorporates a holistic approach, which could serve as an example for innovations for other malignancies.

CML, however, is distinguishable from other malignancies based on the nearly normal survival rates and, above all else, the availability of a molecular marker representing disease activity. Fortunately, survival is nowadays seldom the key issue for patients and the emphasis shifts to the management of AEs, decreasing the disease burden in daily life.

Limitations

Design thinking was an effective method to develop an innovation in a bottom-up approach. However, it is possible that a different product would have been developed if another user-centered method had been used. In addition, our study is limited by the possible bias in patients voluntarily participating

in the development of the innovation. Patients willing to participate in the project could be a more active group, and their needs may not represent the needs of the overall population of patients with CML.

Future Research

The ability of CMyLife to achieve its goal as well as the added value of CMyLife and its features will be the focus of future studies. Pre- and posttest design studies were running at the time of writing this paper. CMyLife will be evaluated on objective criteria (ie, quality of life, cost-effectiveness, adherence rates, etc). In addition, the further development and fine-tuning of CMyLife remains to be a continuous process. During this process, qualitative studies with more focus groups and interviews will be performed.

Enhancing adherence could save health care costs. A Dutch report assessed the potential value of improving both patients' adherence to TKIs and physicians' adherence to guidelines [63]. For each yearly cohort of newly diagnosed patients with CML (160 patients) over a period of 25 years, a cost reduction of approximately US $3.1 million (€2.8 million) and US $1.3 million (€1.2 million) would be possible by improving the patients' adherence and physicians' adherence, respectively, to 100%. For the improvement of every 1% of adherence, cost reduction was established at US $118,460 (€106,560) and US $48,380 (€43.520) per yearly cohort over 25 years, respectively.

Conclusions

The main takeaway of this study is that we have developed a holistic intervention with the aid of active patient participation. We aimed to change the current model of care by using and continuously improving CMyLife. Patients are provided with an easy to understand guideline for the management of CML. In addition, they receive their own laboratory results, including the paramount *BCR-ABL1* values, depicted in a simple yet understandable graph. With the aid of CMyLife, we give patients the tools to monitor their results, interpret these results, and act on them. They are provided with the know-how to consider their results in relation to their personal care process. Patients themselves monitor their treatment and know when it is time to check their values again. After all, physicians proved not to be the most adherent monitors [41]. Of course, the hematologist will not become completely redundant, monitoring is still required, although from a distance. An approximate comparison can be made with a chronic disease such as diabetes mellitus, where patients monitor their own glucose levels and therapy and have, eg, yearly contact with their health care professional. CML would be the first malignancy where patients are able to monitor and manage their disease by themselves.

Historically, CML has been a prototype malignant disease for innovations in medicine: it was the first human cancer in which the causative genetic abnormality was identified (the Philadelphia chromosome) and was the first disease for which a successful targeted anticancer therapy was developed (the TKI imatinib). Perhaps, CML can lead innovation in oncology once again and be the model of change in current health care by redesigning the model of care to be truly centered around the patient, with the patient in full control and ensuring the highest quality and best outcomes.

Acknowledgments

The authors would like to thank all the patients for their contributions to the development of CMyLife.

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

The corresponding author (GE) attests that all listed authors have met authorship criteria and that no others meeting the criteria have been omitted.

Conflicts of Interest

GE, PW, RH, KB, BH, OV, JJ, NB have no conflicts of interests to declare. JJWJ received research support by Novartis and BMS, and honoraria of Abbvie, Novartis, Pfizer and Incyte. He is the president of the Apps for Care and Science Foundation which develops the HematologyApp. This nonprofit foundation is supported by Amgen, Daiichi-Sankyo, Janssen, Incyte, BMS, Servier, Jazz, and Celgene.

Multimedia Appendix 1
Interview guide focus groups patients.
[PDF File (Adobe PDF File), 123 KB - jmir_v22i5e15895_app1.pdf]

Multimedia Appendix 2
Interview guide focus groups hematologists.
[PDF File (Adobe PDF File), 110 KB - jmir_v22i5e15895_app2.pdf]

References

1. Pruitt S. Innovative Care For Chronic Conditions: Building Blocks For Action. Geneva, Switzerland: World Health Organization; 2002.

2. Holman H, Lorig K. Patients as partners in managing chronic disease. Partnership is a prerequisite for effective and efficient health care. Br Med J 2000 Feb 26;320(7234):526-527 [FREE Full text] [doi: 10.1136/bmj.320.7234.526] [Medline: 10688539]
3. Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. J Am Med Assoc 2002 Nov 20;288(19):2469-2475. [doi: 10.1001/jama.288.19.2469] [Medline: 12435261]
4. Pulvirenti M, McMillan J, Lawn S. Empowerment, patient centred care and self-management. Health Expect 2014 Jun;17(3):303-310 [FREE Full text] [doi: 10.1111/j.1369-7625.2011.00757.x] [Medline: 22212306]
5. Bauman AE, Fardy HJ, Harris PG. Getting it right: why bother with patient-centred care? Med J Aust 2003 Sep 1;179(5):253-256. [Medline: 12924973]
6. Jacobson RE. Information Design. Cambridge: The MIT Press; 2000.
7. Brown T. Design thinking. Harv Bus Rev 2008 Jun;86(6):84-92, 141. [Medline: 18605031]
8. Luchs M. A brief introduction to design thinking. In: Luchs M, Swan S, Griffin A, editors. Design Thinking: New Product Development Essentials from the PDMA. New Jersey, USA: John Wiley & Sons Inc; 2015:1-12.
9. Criscitelli T, Goodwin W. Applying human-centered design thinking to enhance safety in the OR. AORN J 2017 Apr;105(4):408-412. [doi: 10.1016/j.aorn.2017.02.004] [Medline: 28336030]
10. Dandara C, Endrenyi L, Kolker E, Hekim N, Steuten LM, Güngör K, et al. Precision Medicine 2.0: The Rise of Glocal Innovation, Superconnectors, and Design Thinking. OMICS 2016 Sep;20(9):493-495. [doi: 10.1089/omi.2016.0129] [Medline: 27631187]
11. Eckman M, Gorski I, Mehta K. Leveraging design thinking to build sustainable mobile health systems. J Med Eng Technol 2016;40(7-8):422-430. [doi: 10.1080/03091902.2016.1218560] [Medline: 27535325]
12. Ferreira F, Song E, Gomes H, Garcia E, Ferreira L. New mindset in scientific method in the health field: Design Thinking. Clinics (Sao Paulo) 2015 Dec;70(12):770-772 [FREE Full text] [doi: 10.6061/clinics/2015(12)01] [Medline: 26735214]
13. Altman M, Huang TT, Breland JY. Design thinking in health care. Prev Chronic Dis 2018 Sep 27;15:E117 [FREE Full text] [doi: 10.5888/pcd15.180128] [Medline: 30264690]
14. Bower H, Björkholm M, Dickman PW, Höglund M, Lambert PC, Andersson TM. Life expectancy of patients with chronic myeloid leukemia approaches the life expectancy of the general population. J Clin Oncol 2016 Aug 20;34(24):2851-2857. [doi: 10.1200/JCO.2015.66.2866] [Medline: 27325849]
15. Huang X, Cortes J, Kantarjian H. Estimations of the increasing prevalence and plateau prevalence of chronic myeloid leukemia in the era of tyrosine kinase inhibitor therapy. Cancer 2012 Jun 15;118(12):3123-3127 [FREE Full text] [doi: 10.1002/cncr.26679] [Medline: 22294282]
16. Baccarani M, Deininger MW, Rosti G, Hochhaus A, Soverini S, Apperley JF, et al. European LeukemiaNet recommendations for the management of chronic myeloid leukemia: 2013. Blood 2013 Aug 8;122(6):872-884 [FREE Full text] [doi: 10.1182/blood-2013-05-501569] [Medline: 23803709]
17. O'Brien S, Berman E, Borghaei H, Deangelo DJ, Devetten MP, Devine S, National Comprehensive Cancer Network. NCCN clinical practice guidelines in oncology: chronic myelogenous leukemia. J Natl Compr Canc Netw 2009 Oct;7(9):984-1023. [doi: 10.6004/jnccn.2009.0065] [Medline: 19878641]
18. Goldberg SL. Monitoring chronic myeloid leukemia in the real world: gaps and opportunities. Clin Lymphoma Myeloma Leuk 2015 Dec;15(12):711-714. [doi: 10.1016/j.clml.2015.08.088] [Medline: 26433907]
19. Guérin A, Chen L, Dea K, Wu EQ, Goldberg SL. Association between regular molecular monitoring and tyrosine kinase inhibitor therapy adherence in chronic myelogenous leukemia in the chronic phase. Curr Med Res Opin 2014 Jul;30(7):1345-1352. [doi: 10.1185/03007995.2014.904281] [Medline: 24640967]
20. Saleh MN, Haislip S, Sharpe J, Hess T, Gilmore J, Jackson J, et al. Assessment of treatment and monitoring patterns and subsequent outcomes among patients with chronic myeloid leukemia treated with imatinib in a community setting. Curr Med Res Opin 2014 Apr;30(4):529-536. [doi: 10.1185/03007995.2013.858621] [Medline: 24156689]
21. Goldberg SL, Chen L, Guerin A, Macalalad AR, Liu N, Kaminsky M, et al. Association between molecular monitoring and long-term outcomes in chronic myelogenous leukemia patients treated with first line imatinib. Curr Med Res Opin 2013 Sep;29(9):1075-1082. [doi: 10.1185/03007995.2013.812034] [Medline: 23738923]
22. Marin D, Bazeos A, Mahon F, Eliasson L, Milojkovic D, Bua M, et al. Adherence is the critical factor for achieving molecular responses in patients with chronic myeloid leukemia who achieve complete cytogenetic responses on imatinib. J Clin Oncol 2010 May 10;28(14):2381-2388 [FREE Full text] [doi: 10.1200/JCO.2009.26.3087] [Medline: 20385986]
23. Noens L, van Lierde MA, de Bock R, Verhoef G, Zachée P, Berneman Z, et al. Prevalence, determinants, and outcomes of nonadherence to imatinib therapy in patients with chronic myeloid leukemia: the ADAGIO study. Blood 2009 May 28;113(22):5401-5411. [doi: 10.1182/blood-2008-12-196543] [Medline: 19349618]
24. Latremouille-Viau D, Guerin A, Gagnon-Sanschagrin P, Dea K, Cohen BG, Joseph GJ. Health care resource utilization and costs in patients with chronic myeloid leukemia with better adherence to tyrosine kinase inhibitors and increased molecular monitoring frequency. J Manag Care Spec Pharm 2017 Feb;23(2):214-224 [FREE Full text] [doi: 10.18553/jmcp.2017.23.2.214] [Medline: 28125373]

25. Efficace F, Baccarani M, Breccia M, Saussele S, Abel G, Caocci G, et al. International development of an EORTC questionnaire for assessing health-related quality of life in chronic myeloid leukemia patients: the EORTC QLQ-CML24. Qual Life Res 2014 Apr;23(3):825-836. [doi: 10.1007/s11136-013-0523-5] [Medline: 24026634]
26. Steegmann JL, Baccarani M, Breccia M, Casado LF, García-Gutiérrez V, Hochhaus A, et al. European LeukemiaNet recommendations for the management and avoidance of adverse events of treatment in chronic myeloid leukaemia. Leukemia 2016 Aug;30(8):1648-1671 [FREE Full text] [doi: 10.1038/leu.2016.104] [Medline: 27121688]
27. Efficace F, Rosti G, Aaronson N, Cottone F, Angelucci E, Molica S, et al. Patient- versus physician-reporting of symptoms and health status in chronic myeloid leukemia. Haematologica 2014 Apr;99(4):788-793 [FREE Full text] [doi: 10.3324/haematol.2013.093724] [Medline: 24241488]
28. Hasso Plattner Institute of Design at Stanford University. Stanford d.school. URL: https://dschool.stanford.edu/ [accessed 2018-01-03] [WebCite Cache ID 6wC3Ai4Vz]
29. Liu J. Medium – Get smarter about what matters to you. 2016. Visualizing the 4 Essentials of Design Thinking Medium URL: https://medium.com/good-design/visualizing-the-4-essentials-of-design-thinking-17fe5c191c22 [accessed 2018-02-19] [WebCite Cache ID 6xLWlLZLz]
30. Dam R, Siang T. Interaction Design Foundation. 2017. Stage 2 in the Design Thinking Process: Define the Problem and Interpret the Results URL: https://www.interaction-design.org/literature/article/stage-2-in-the-design-thinking-process-define-the-problem-and-interpret-the-results [accessed 2018-01-03] [WebCite Cache ID 6wC2HUmdO]
31. IDEO. The Field Guide To Human-centered Design: Design Kit. Canada: IDEO; 2015.
32. Ohno T. Toyota Production System: Beyond Large-scale Production. Boca Raton, Florida: CRC Press; 1978.
33. Vianna M, Vianna Y, Adler I, Lucena B, Russo B. Design Thinking: Business Innovation. Rio de Janeiro: MJV Press; 2013.
34. Saußele S, Richter J, Hochhaus A, Mahon FX. The concept of treatment-free remission in chronic myeloid leukemia. Leukemia 2016 Aug;30(8):1638-1647 [FREE Full text] [doi: 10.1038/leu.2016.115] [Medline: 27133824]
35. Hughes TP, Ross DM. Moving treatment-free remission into mainstream clinical practice in CML. Blood 2016 Jul 7;128(1):17-23. [doi: 10.1182/blood-2016-01-694265] [Medline: 27013442]
36. Saglio G, Sharf G, Almeida A, Bogdanovic A, Bombaci F, Čugurović J, et al. Considerations for treatment-free remission in patients with chronic myeloid leukemia: a joint patient-physician perspective. Clin Lymphoma Myeloma Leuk 2018 Jun;18(6):375-379 [FREE Full text] [doi: 10.1016/j.clml.2018.04.005] [Medline: 29753691]
37. Patients Know Best. URL: https://www.patientsknowbest.com/ [accessed 2018-01-03] [WebCite Cache ID 6wAlmX2lV]
38. MedApp. URL: https://www.medapp.nu/en/ [accessed 2018-03-01] [WebCite Cache ID 6wC2qZEpS]
39. Janssen J, Cornelissen J, Posthuma E, Falkenburg J, Biemond BJ, Bos G, et al. Richtlijnen voor de behandeling van chronische myeloide leukemie anno 2014. Ned Tijdschr Hematol 2014;11(5):- [FREE Full text]
40. Jackson J, Chen L, Sail K, Ericson S, Haislip S, Sharpe J, et al. Monitoring Patterns and Subsequent Outcomes Among Patients with Philadelphia Chromosome Positive Chronic Myeloid Leukemia (Ph+ CML) Treated with Imatinib in a Community Setting. Blood 2012;120(21):2066. [doi: 10.1182/blood.V120.21.2066.2066]
41. Geelen IG, Thielen N, Janssen JJ, Hoogendoorn M, Roosma TJ, Willemsen SP, et al. Treatment outcome in a population-based, 'real-world' cohort of patients with chronic myeloid leukemia. Haematologica 2017 Nov;102(11):1842-1849 [FREE Full text] [doi: 10.3324/haematol.2017.174953] [Medline: 28860339]
42. Geelen IG, Thielen N, Janssen JJ, Hoogendoorn M, Roosma TJ, Willemsen SP, et al. Impact of hospital experience on the quality of tyrosine kinase inhibitor response monitoring and consequence for chronic myeloid leukemia patient survival. Haematologica 2017 Dec;102(12):e486-e489 [FREE Full text] [doi: 10.3324/haematol.2017.175265] [Medline: 28860340]
43. Mahon F, Réa D, Guilhot J, Guilhot F, Huguet F, Nicolini F, Intergroupe Français des Leucémies Myéloïdes Chroniques. Discontinuation of imatinib in patients with chronic myeloid leukaemia who have maintained complete molecular remission for at least 2 years: the prospective, multicentre Stop Imatinib (STIM) trial. Lancet Oncol 2010 Nov;11(11):1029-1035. [doi: 10.1016/S1470-2045(10)70233-3] [Medline: 20965785]
44. Ross DM, Branford S, Seymour JF, Schwarer AP, Arthur C, Yeung DT, et al. Safety and efficacy of imatinib cessation for CML patients with stable undetectable minimal residual disease: results from the TWISTER study. Blood 2013 Jul 25;122(4):515-522. [doi: 10.1182/blood-2013-02-483750] [Medline: 23704092]
45. Saussele S, Richter J, Guilhot J, Gruber FX, Hjorth-Hansen H, Almeida A, EURO-SKI investigators. Discontinuation of tyrosine kinase inhibitor therapy in chronic myeloid leukaemia (EURO-SKI): a prespecified interim analysis of a prospective, multicentre, non-randomised, trial. Lancet Oncol 2018 Jun;19(6):747-757. [doi: 10.1016/S1470-2045(18)30192-X] [Medline: 29735299]
46. Breccia M, Efficace F, Alimena G. Adherence to treatment is a complex and multifaceted issue that can substantially alter the outcome of chronic myeloid leukemia patients treated with tyrosine kinase inhibitors. Leuk Res 2012 Jul;36(7):804-805. [doi: 10.1016/j.leukres.2012.02.032] [Medline: 22459331]
47. Breccia M, Efficace F, Sica S, Abruzzese E, Cedrone M, Turri D, et al. Adherence and future discontinuation of tyrosine kinase inhibitors in chronic phase chronic myeloid leukemia. A patient-based survey on 1133 patients. Leuk Res 2015 Oct;39(10):1055-1059. [doi: 10.1016/j.leukres.2015.07.004] [Medline: 26282944]

48. Efficace F, Baccarani M, Rosti G, Cottone F, Castagnetti F, Breccia M, et al. Investigating factors associated with adherence behaviour in patients with chronic myeloid leukemia: an observational patient-centered outcome study. Br J Cancer 2012 Sep 4;107(6):904-909 [FREE Full text] [doi: 10.1038/bjc.2012.348] [Medline: 22871884]
49. Eliasson L, Clifford S, Barber N, Marin D. Exploring chronic myeloid leukemia patients' reasons for not adhering to the oral anticancer drug imatinib as prescribed. Leuk Res 2011 May;35(5):626-630. [doi: 10.1016/j.leukres.2010.10.017] [Medline: 21095002]
50. Unnikrishnan R, Veeraiah S, Mani S, Rajendranath R, Rajaraman S, Elangovan GS, et al. Comprehensive evaluation of adherence to therapy, its associations, and its implications in patients with chronic myeloid leukemia receiving imatinib. Clin Lymphoma Myeloma Leuk 2016 Jun;16(6):366-71.e3. [doi: 10.1016/j.clml.2016.02.040] [Medline: 27052853]
51. Rychter A, Jerzmanowski P, Hołub A, Specht-Szwoch Z, Kalinowska V, Tęgowska U, et al. Treatment adherence in chronic myeloid leukaemia patients receiving tyrosine kinase inhibitors. Med Oncol 2017 Jun;34(6):104 [FREE Full text] [doi: 10.1007/s12032-017-0958-6] [Medline: 28444623]
52. Ibrahim AR, Eliasson L, Apperley JF, Milojkovic D, Bua M, Szydlo R, et al. Poor adherence is the main reason for loss of CCyR and imatinib failure for chronic myeloid leukemia patients on long-term therapy. Blood 2011 Apr 7;117(14):3733-3736 [FREE Full text] [doi: 10.1182/blood-2010-10-309807] [Medline: 21346253]
53. Kekäle M, Talvensaari K, Koskenvesa P, Porkka K, Airaksinen M. Chronic myeloid leukemia patients' adherence to peroral tyrosine kinase inhibitors compared with adherence as estimated by their physicians. Patient Prefer Adherence 2014;8:1619-1627 [FREE Full text] [doi: 10.2147/PPA.S70712] [Medline: 25473270]
54. Stefanovic S, Wallwiener M, Karic U, Domschke C, Katic L, Taran F, et al. Patient-reported outcomes (PRO) focused on adverse events (PRO-AEs) in adjuvant and metastatic breast cancer: clinical and translational implications. Support Care Cancer 2017 Feb;25(2):549-558. [doi: 10.1007/s00520-016-3437-2] [Medline: 27747478]
55. Velikova G, Booth L, Smith AB, Brown PM, Lynch P, Brown JM, et al. Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. J Clin Oncol 2004 Feb 15;22(4):714-724. [doi: 10.1200/JCO.2004.06.078] [Medline: 14966096]
56. de Marchi F, Medeot M, Fanin R, Tiribelli M. How could patient reported outcomes improve patient management in chronic myeloid leukemia? Expert Rev Hematol 2017 Jan;10(1):9-14. [doi: 10.1080/17474086.2017.1262758] [Medline: 27858461]
57. Kuijpers W, Groen WG, Oldenburg HS, Wouters MW, Aaronson NK, van Harten WH. eHealth for breast cancer survivors: use, feasibility and impact of an interactive portal. JMIR Cancer 2016 May 10;2(1):e3 [FREE Full text] [doi: 10.2196/cancer.5456] [Medline: 28410178]
58. Groen WG, Kuijpers W, Oldenburg HS, Wouters MW, Aaronson NK, van Harten WH. Supporting lung cancer patients with an interactive patient portal: feasibility study. JMIR Cancer 2017 Aug 8;3(2):e10 [FREE Full text] [doi: 10.2196/cancer.7443] [Medline: 28790025]
59. van Bruinessen IR, van Weel-Baumgarten EM, Gouw H, Zijlstra JM, van Dulmen S. An integrated process and outcome evaluation of a web-based communication tool for patients with malignant lymphoma: randomized controlled trial. J Med Internet Res 2016 Jul 29;18(7):e206 [FREE Full text] [doi: 10.2196/jmir.5877] [Medline: 27473173]
60. van der Hout A, van Uden-Kraan CF, Holtmaat K, Jansen F, Lissenberg-Witte BI, Nieuwenhuijzen GA, et al. Role of eHealth application Oncokompas in supporting self-management of symptoms and health-related quality of life in cancer survivors: a randomised, controlled trial. Lancet Oncol 2020 Jan;21(1):80-94. [doi: 10.1016/S1470-2045(19)30675-8] [Medline: 31838009]
61. Kaal SE, Husson O, van Dartel F, Hermans K, Jansen R, Manten-Horst E, et al. Online support community for adolescents and young adults (AYAs) with cancer: user statistics, evaluation, and content analysis. Patient Prefer Adherence 2018;12:2615-2622 [FREE Full text] [doi: 10.2147/PPA.S171892] [Medline: 30584285]
62. Bender JL, Jimenez-Marroquin MC, Ferris LE, Katz J, Jadad AR. Online communities for breast cancer survivors: a review and analysis of their characteristics and levels of use. Support Care Cancer 2013 May;21(5):1253-1263. [doi: 10.1007/s00520-012-1655-9] [Medline: 23179491]
63. Ector GI, Govers TM, Westerweel PE, Grutters JP, Blijlevens NM. The potential health gain and cost savings of improving adherence in chronic myeloid leukemia. Leuk Lymphoma 2019 Jun;60(6):1485-1492. [doi: 10.1080/10428194.2018.1535113] [Medline: 30668187]

Abbreviations

AE: adverse event
CML: chronic myeloid leukemia
eHealth: electronic health
PRO: patient-reported outcome
TFR: treatment-free remission
TKI: tyrosine kinase inhibitor

Edited by G Eysenbach; submitted 16.08.19; peer-reviewed by H Oh, A Kouroubali, E Seto; comments to author 02.11.19; revised version received 11.01.20; accepted 07.02.20; published 15.05.20.

Please cite as:
Ector GICG, Westerweel PE, Hermens RPMG, Braspenning KAE, Heeren BCM, Vinck OMF, de Jong JJM, Janssen JJWM, Blijlevens NMA
The Development of a Web-Based, Patient-Centered Intervention for Patients With Chronic Myeloid Leukemia (CMyLife): Design Thinking Development Approach
J Med Internet Res 2020;22(5):e15895
URL: https://www.jmir.org/2020/5/e15895
doi:10.2196/15895
PMID:32412424

Original Paper

Mapping Canadian Men's Recent and Intended Health Behavior Changes Through the Don't Change Much Electronic Health Program

John L Oliffe[1,2], RN, MEd, PhD; Nick Black[3], BBus, MA; Jeffrey Yiu[3], BSc; Ryan K Flannigan[4,5], MD; Donald R McCreary[6], PhD; S Larry Goldenberg[4], CM, OBC, MD, FRCSC

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

[2]Department of Nursing, University of Melbourne, Melbourne, Australia

[3]Intensions Consulting Inc, Vancouver, BC, Canada

[4]Department of Urologic Sciences, University of British Columbia, Vancouver, BC, Canada

[5]Department of Urology, Weill Cornell Medicine, New York, NY, United States

[6]Department of Psychology, Brock University, St. Catharines, ON, Canada

Corresponding Author:
John L Oliffe, RN, MEd, PhD
School of Nursing
Faculty of Applied Science
University of British Columbia
T201-2211 Wesbrook Mall
Vancouver, BC, V6T 2B5
Canada
Phone: 1 604 822 7638
Email: john.oliffe@ubc.ca

Abstract

Background: Although evaluation studies confirm the strong potential of men's electronic health (eHealth) programs, there have been calls to more fully understand acceptability, engagement, and behavior change to guide future work. Relatedly, mapping of behavior changes using health promotion theories including the transtheoretical model (or stages of change) has been recommended to build a translatable empirical base to advance design and evaluation considerations for men's eHealth programs.

Objective: This study aimed to use a benchmark sample as a reference group to map the recent and intended health behavior changes in Canadian men who use the *Don't Change Much* (DCM) eHealth program. The hypothesis being tested was that increased exposure to DCM would be positively associated with men's recent and intended health behavior changes.

Methods: DCM users (n=863) were sampled for demographic data and self-reported recent and intended health behavior changes. Respondents also reported their usage (frequency and duration) for each of the 3 DCM components (web, newsletter, and social media) and were allocated to limited exposure (257/863, 29.8%), low exposure (431/863, 49.9%), and high exposure (175/863, 20.3%) subgroups. A benchmark sample (n=2000), comprising respondents who had not accessed DCM provided a reference group. Bivariate analysis of recent and intended health behavior changes and DCM exposure levels were used to compute the strength of association between the independent variables (exposure levels) and the 10 categorical dependent variables (recent and intended health behavior changes). Binary logistic regression models were computed for each of the 10 recent and intended health behavior changes. Linear regression was used to model the association between the number of recent and intended changes and the level of exposure to DCM.

Results: Compared with the benchmark reference group, DCM high-exposure respondents had significantly increased odds for 9 of the 10 health behavior changes, with the largest effect size observed for Changed diet or Improved eating habits (odds ratio [OR] 5.628, 95% CI 3.932-8.055). High-exposure respondents also had significantly increased odds for 9 intended health changes, with the largest effect sizes observed for Reduce stress level (OR 4.282, 95% CI 3.086-5.941). Moderate effect size (goodness of fit) was observed for increased total number of recent ($F_{12,2850}$=25.52; *P*.001; adjusted R^2=.093) and intended health behavior changes ($F_{12,2850}$=36.30; *P*.001; adjusted R^2=.129) among high-exposure respondents.

Conclusions: DCM respondents contrasted the predominately precontemplative benchmark sample mapping across the contemplative, preparation, and action stages of the transtheoretical health behavior change model. Almost 10% of variation in the recent and 13% of variation in the intended health behavior changes can be explained by DCM exposure and demographic factors, indicating the acceptability of this men's eHealth resource.

(J Med Internet Res 2020;22(5):e16174) doi:10.2196/16174

KEYWORDS

men's e-Health; men's health behavior change; men's health promotion

Introduction

The case for men's health is often articulated through sex differences research wherein lower life expectancy in men (compared with that of women) is connected to their overall poor self-health practices, including estrangement from professional *in person* health care services [1,2]. The major mortality causes accounting for men's reduced life expectancy include cardiovascular disease, suicide, motor vehicle accidents, liver failure (most often due to alcohol overuse), and infectious diseases (most often HIV) [3]. These (and many other) mortality causes and contributors to morbidities are seemingly amenable to prevention-based interventions, and by virtue of that, tailored health promotion programs have surfaced to garner men's health behavior changes [4-6]. The platforms and mechanisms for promoting men's health, although diverse, have grown exponentially in the electronic health (eHealth) sector over the last two decades [7,8]. That said, empirical insights to the acceptability, engagement levels, and behavior changes reaped through these well-intended men's eHealth programs, while promising, are emergent and drawn from diverse study designs [7,9]. The aim of this study was to use a benchmark sample as a reference group to map the recent and intended health behavior changes in Canadian men who use the *Don't Change Much* (DCM) eHealth program [10]. The hypothesis being tested was that increased DCM exposure levels would be positively associated with men's recent and intended health behavior changes.

Men's lifestyles have attracted health promotion research describing specific risk factors and a range of potential remedies. In terms of risk, unhealthy diets, alcohol overuse, smoking, and sedentary lifestyles have featured as issues warranting tailored behavior change interventions [11]. There is also diversity in the predisposition to the aforementioned behavior risks whereby inequities within social determinants of health (ie, income, employment, education) result in disadvantage to some male subgroups both in terms of their knowledge levels and access to health care services [12,13]. Within this context, men's health risks and the potential for behavior change emerge as somewhat relative, deeply reliant on resources being freely available and easily accessible. By lessening structural barriers and mobilizing men's strength-based efforts for optimizing their health, important behavior changes can occur for men [13]. In this regard, men's eHealth resources have great potential to improve access and meet many health promotion needs for men.

Men's eHealth programs have grown significantly to deliver diverse information and services across ever expansive platforms (ie, web, social media, and email). The wide variety of men's eHealth programs include interventions tailored to address weight loss [14], smoking cessation [15], prenatal health education [16], fathering [17], depression management [7,18], sexual health [19-24], and prostate cancer [25]. Although there are claims that eHealth resources can engage diverse subgroups of men, satiating their preferences for anonymity and self-directed health help-seeking, two significant issues prevail. The first relates to the varied conceptualizations and approaches to evaluating needs analyses (acceptability) and end-user engagement, as well as the lack of conclusive empirical evidence regarding associations between men's eHealth programs and behavior change. For example, needs analysis of men transitioning to fatherhood by Da Costa et al [17] confirmed substantial interest among new and expectant fathers for using internet-delivered strategies to promote their mental health and prepare for parenthood. Within this and similar needs analyses work, the insights drawn from potential end users have affirmed the acceptability of men's eHealth programs in specific contexts and informed tailored content and targeted delivery of the interventions. Evaluations of newly launched and established eHealth programs have tended to focus on men's engagement or linkages to behavior change [16,22]. Examples include a single-group, repeated measures design evaluating the applicability of Man Central (a web and mobile phone intervention for men with depression), which showed significant improvements in depression symptoms, depression risk, externalizing symptoms, and work and social functioning among end users [7]. Pre-post eHealth evaluation studies reported significant improvements in depressive symptoms among a community sample of overweight and obese men with depression [14], whereas Bottorff et al [15] highlighted tobacco reduction and/or cessation for 66% (43) of men who used the QuitNowMen resource. Klein et al's [24] quasi-experimental, 2-arm study evaluated *Real Talk* (an eHealth harm reduction intervention targeting black men who have sex with men) reported end user's HIV knowledge gains (although there were no significant differences between *Real Talk* and the control participant's actual condom use or other risk reduction strategies). Although the aforementioned and many other men's eHealth evaluation studies have been limited by small samples, attrition, and/or a lack of control groups, the results are encouraging [25,26]. In sum, acceptability and evaluation studies confirm the strong potential of men's eHealth programs amid calls for more evidence to efficiently guide future work and confidently claim sustained health behavior change effects.

The second issue relates to incorporating theory to map men's health behavior changes and their relationships to tailored eHealth programs. Simoni et al [26] argued the need for health promotion theory in men's eHealth to build a translatable

empirical base and advance the design and evaluation of gender-sensitized interventions. Among many health promotion theories, the transtheoretical (or stages of change) model [27,28] has guided eHealth program designs and helped contextualize barriers and facilitators to men's health behavior changes in a range of contexts [29-31]. Comprising precontemplative, contemplative, preparation, action, and maintenance, these interconnected and recursive stages of change characterize the transtheoretical model, underpinned by processes reliant on men's access, self-efficacy and recognition of, and commitment to the benefits of sustained modifications. Program design, acceptability, engagement, and behavior change evaluations of men's eHealth resources can, and many argue should, interface with the transtheoretical model (or similar) to map men's progress and adjust intervention content and/or delivery accordingly [29-31]. In line with this recommendation, and reflecting our commitment to fully integrating the transtheoretical model, this study adds to methodological approaches by using a benchmark sample as a reference group to map men's eHealth engagement and behavior change, in making recommendations for DCM, and, more broadly, the burgeoning field of men's eHealth.

Methods

Overview

Following university ethics approvals, demographic and self-reported recent and intended health behavior change data were collected via survey questionnaires from two Canadian male cohorts: (1) benchmark (reference group), comprising respondents who had not accessed DCM, and (2) DCM users. Data collection details and cross-sectional findings from the benchmark sample have been reported elsewhere [6,32]. Briefly, the 5083 respondents recruited via web-based panel provider to complete a Canadian men's health survey were reduced to 2000, stratified by age and location to be close to the most recent Canadian census data. The 15-min web benchmark survey was administered from April 20, 2017, to April 28, 2017.

The second cohort comprised DCM users recruited via the DCM website and email newsletter recipient list, and these respondents were incentivized with the option to enter a prize draw to win Can $500 (US $377.60) cash. By way of background, the DCM eHealth program was purpose-built in 2014 by the Canadian Men's Health Foundation, a national nonprofit organization, with the goal of inspiring men and their families to lead healthier lives. Reliable information and easily accessible tips are developed based on focus group interviews with Canadian men and reviews of the men's health literature, and by drawing from the expertise of global thought leaders. The DCM information shared through testimonials, text, video, and audio is brief, often times humorous, and always easy to access with a focus on practical strategies to improve diet, exercise, sleep, and stress management as well as reduce alcohol use and/or smoking. A range of strategies are offered in point form recognizing that end users are diverse in their contexts, needs, and alignments to the transtheoretical stages of change (ie, precontemplative, contemplative, preparation, action, and maintenance). In addition to providing a framework for evaluating DCM, the program content was developed drawing on the transtheoretical model. Specifically, DCM's strength-based approach and wide-reaching materials were purpose-built to engage men at diverse levels of readiness to change and progress points. In essence, the interconnected stages characterizing the transtheoretical model guided the DCM content design to work with men at whatever point they were at, to advance their self-health.

The DCM demographic and survey questionnaire data were collected between January 1, 2018, and March 31, 2018; the data collection tools were identical to those of the benchmark survey but also included questions about the respondent's usage (duration and frequency) of the three DCM components (web, newsletter, and social media). Of the 1743 respondents who went to the DCM survey introduction page, 94.8% (1653) opted in. This sample was reduced to 1034 Canadian male DCM users by removing incomplete surveys (n=459), female respondents (n=90), speeding and/or straight lining responses (n=60), respondents from outside of Canada (n=7), and those under the age of 19 years (n=3). Some demographics were asked at the end of the survey (education, household income, and ethnicity), making it impossible to accurately evaluate and report the demographic characteristics for the 459 incomplete surveys. Speeding was assigned to respondents who completed the survey in 5 min or less (less than one-third of the median completion time), and straight lining comprised respondents inputting the same numerical response to all Likert items on 2 or more consecutive survey pages. The final sample of 863 DCM users was obtained through listwise removal of an additional 171 respondents who answered *not sure* to questions regarding their use (duration and frequency) of the 3 DCM components, as a numerical code could not be assigned to classify those respondents' DCM exposure level.

Measures

Demographic data, including age, employment, living arrangements (lives alone; children younger than 19 years living at home), education, visible minority, sexual orientation, and household income (before taxes) were collected. *Visible minority* is the terminology used by Statistics Canada, and based on that classification system, we used the term in our postcoding. In addition to being a Canadian resident, respondents were asked: *Do you belong to any distinct ethnic or cultural group?* with response options: *Yes (please specify)* or *No (Prefer not to say)*. Responses from those who answered *Yes (please specify)* were postcoded using Statistics Canada 2016 Census definitions. Regards sexual orientation, respondents were asked the following: *Do you consider yourself to be: Select one: Heterosexual or straight, Gay or lesbian, Bisexual, Not sure or questioning, or Other (please specify)*. Responses were coded as a dichotomous variable (0=heterosexual or straight, 1=other).

Respondents in both cohorts were also asked about the recent and intended changes to improve their health. The first question asked, *In the past 12-months, have you made any changes that would improve your health?* inviting respondents to select all that applied from the following 10 items: (1) changed diet or improved eating habits, (2) made an effort to sit less and walk more, (3) increased exercise, sports or physical activity, (4) I haven't made any changes, (5) drink less alcohol, (6) had a

routine check-up or visit to doctor, (7) improved consistent sleep quality, (8) lost weight, (9) reduced stress level, and (10) quit or reduced smoking. Intended changes were collected through soliciting responses to the same 10 options, with the stem question, *In the next month (30-days), do you intend to make any changes that would improve your health? Select all that apply*.

The 863 DCM respondents had access to three DCM eHealth components: (1) the website, (2) email newsletter, and (3) social media accounts (Facebook, Instagram, and Twitter). Respondents checked yes, no or not sure options for each of the three DCM components in answering the question, *Have you ever used, or subscribed to, any of the following DCM resources?* Respondents who checked *Not sure* responses were excluded. *No* responses were allocated a zero value and assigned to the limited exposure subgroup. *Yes* responses were used to subsequently populate DCM items (website, newsletter, and social media) in asking 2 additional questions to score each respondent's DCM usage (duration and frequency) and calculating exposure levels [33]. The work of Quinn and Chaudoir [33] guided the assignment of numerical codes to participant's single-item frequency and duration responses for the DCM components that they had used.

1. *When did you FIRST use, or subscribe to, the following resources?* Response options: never before, in the past month, 1-6 months ago, 7-12 months ago, 13-24 months ago, more than 2 years ago, and not sure. Participants checked the duration level based on their first use for the DCM resources. The categorical measure was converted to a continuous measure representing the duration as follows: *never before*=0.0, *in the past month*=0.5, *1-6 months ago*=3.5, *7-12 months ago*=9.5, *13-24 months ago*=18.5, *more than 2 years ago*=24.0, and not sure (excluded).
2. *How often do you use or access the following resources?* Response options were as follows: several times a day, once a day, several times a week, once a week, several times a month, once a month, several times a year, once a year, less often, do not use, and not sure. To calculate frequency, the checked categorical measure was converted into a continuous measure: *several times a day*=1095, *once a day*=365, *several times a week*=156, once a week=52, several times a month=36, *once a month*=12, *several times a year*=3, *once a year*=1, *less often* =0.5, *do not use*=0, and not sure (excluded).

DCM user classifications were based on the sum of the product of duration and frequency for each of the three DCM components, to calculate respondent exposure scores. Duration responses were converted to numerical scores that represented the number of months since accessing DCM. Frequency responses were converted to numerical scores proportional to the number of times each DCM component was used. To illustrate, respondents who completed the survey questionnaire but indicated a duration of *Never before* and a frequency of *Do not use* for all three DCM components received a summed score of 0 and were classified to the limited-exposure group (257/863, 29.8%). A user who engaged one DCM resource for a duration of *In the past month* (0.5) at a frequency of *Less often* (0.5) received a total summed score of 0.25 and was classified to the low-exposure subgroup (431/863, 49.9%; range 0.25-680). A respondent who used two DCM programs, both for a duration of *7-12 months ago* (9.5) at a frequency of *Several times a month* (36), received a total summed score of 684 and was classified as high exposure (175/863, 20.3%; range 684-9048). Through these classifications, three DCM user subgroups, limited, low, and high exposure, were delineated.

Data Analysis

Bivariate analyses were conducted to assess the magnitude of differences in demographic factors between benchmark respondents and DCM users. The chi-square test was used to assess whether there was an association between the two categorical variables. Cohen *d* was used to calculate effect size. When the means of three or more independent groups were compared, Cohen *d* was obtained by computing partial eta-squared (another type of effect size) and then converting partial eta-squared to Cohen *d* using formulae in the study by Cohen [34]. Cramer's V was used to calculate an effect size for the strength of association between the two categorical variables, with values ranging from 0 to 1 (inclusive) [35].

Regression analysis included logistic regression to model the association between the dependent variables recent and intended health behavior changes, and independent variable, level of exposure to DCM. R-squared values were used to measure the proportion of variance in a dependent variable that can be explained by the independent variable in the regression models [34]. For models analyzing recent changes, the dependent variable was whether the user made the specified behavior change (eg, *Increased exercise, sports or physical activity*) in the past 12 months or not. Similarly, for models analyzing intended changes, the dependent variable was whether the user intended to make the specified behavior change (eg, *Change diet or improve eating habits*) in the next month (30 days) or not. Linear regression was used to model the association between the number of recent and intended changes and the level of exposure to DCM. In total, two linear regression models were computed, with the first model having the number of recent changes as the dependent variable, and the second model having the number of intended changes as the dependent variable. The DCM users were classified into 1 of 3 categories, limited, low, or high exposure. Benchmark respondents, classified as no exposure, were the reference group. All logistic and linear regression models controlled for the following covariates: age, employment, living arrangements (lives with partner; children younger than 19 years living at home), education, visible minority, sexual orientation, and household income (before taxes). Odds ratios are effect sizes and were used to indicate the strength of association between predictor variables and dichotomous outcome variables [36]. Variance inflation factors (VIFs) were computed as a collinearity diagnostic check.

Results

Benchmark respondents had not used DCM and served as a reference group to assess the effect of DCM program exposures on recent and intended health behavior changes. Assessment of the magnitude of differences in demographic factors between

benchmark respondents and DCM users revealed small to negligible effect sizes for the 8 demographic factors (see Table 1). Most respondents in both cohorts reported being employed, living with a partner, not living with children younger than 19 years, not having graduated from university, and identifying as heterosexual. There were no statistically significant differences in age between the two cohorts. Compared with benchmark respondents, a higher proportion of DCM users reported having a household income of Can $120,000 (US $92,307.69) or more, although household income and the two cohorts were weakly associated as a whole.

For the DCM group, all recent health behavior changes had significant associations with increased levels of exposure to DCM. Moderate effect sizes were observed in *Changed diet or improved eating habits* (χ^2_3=210; *P*<.001; Cramer's V=0.271) and *Made an effort to sit less and walk more* (χ^2_3=167; *P*<.001; Cramer's V=0.242). All intended health behavior changes had significant associations with increased DCM exposure. Moderate effect sizes were observed for the following seven intended health behavior changes: (1) *Improve consistent sleep quality* (χ^2_3=207; *P*<.001; Cramer's V=0.269), (2) *Changed diet or improved eating habits* (χ^2_3=181; *P*<.001; Cramer's V=0.252), (3) *Increase exercise, sports or physical activity* (χ^2_3=169; *P*<.001; Cramer's V=0.243), (4) *Make an effort to sit less and walk more* (χ^2_3=168; *P*<.001; Cramer's V=0.242), (5) *I don't intend to make any changes* (χ^2_3=148; *P*<.001; Cramer's V=0.227), (6) *Reduce stress levels* (χ^2_3=129; *P*<.001; Cramer's V=0.212), and (7) *Lose weight* (χ^2_3=119; *P*<.001; Cramer's V=0.205). There was a significant association between the total number of recent health changes and increased DCM exposure levels (ranging no exposure 2.39 to 4.23 high exposure), with a large effect size observed (*P*<.001; *d*=0.818). There was a significant association between the total number of intended health changes and increased DCM exposure levels (ranging no exposure 2.09 to 3.92 high exposure), with a very large effect size observed (*P*<.001; *d*=1.018). Following Bonferroni-adjusted Dunn pairwise tests, with the no-exposure benchmark cohort, all recent and intended changes in the limited-, low-, and high-exposure DCM subgroups were statistically significant (*P*<.001). Recent changes were also statistically significant for DCM limited- and high-exposure subgroups (*P*<.001), as well as for the DCM low- and high-exposure subgroups (*P*=.01).

Table 1. Benchmark versus Don't Change Much users' sample profile.

Demographics and baseline characteristics	Benchmark (n=2000)	DCM[a] users (n=863)	Chi square (*df*)	*P* value	Cramer's V
Age, mean (SD)	46.99 (15.67)	47.34 (11.96)	–0.646[b]	.52	.024[c]
Employed, n (%)	N/A[d]	N/A	51.319 (1)	<.001	0.134
Yes	1307 (65.4)	680 (78.8)	N/A	N/A	N/A
No	693 (34.6)	183 (21.2)	N/A	N/A	N/A
Partner living with respondent, n (%)	N/A	N/A	17.947 (1)	<.001	0.079
Yes	1210 (60.5)	594 (68.8)	N/A	N/A	N/A
No	790 (39.5)	269 (31.2)	N/A	N/A	N/A
Children <19 years living with respondent, n (%)	N/A	N/A	80.292 (1)	<.001	0.167
Yes	441 (22.0)	330 (38.2)	N/A	N/A	N/A
No	1559 (78.0)	533 (61.8)	N/A	N/A	N/A
Highest level of education, n (%)	N/A	N/A	9.234 (1)	.002	0.057
Graduated university	823 (41.2)	408 (47.3)	N/A	N/A	N/A
Other	1177 (58.8)	455 (52.7)	N/A	N/A	N/A
Visible minority, n (%)			.128 (1)	.72	0.007
Yes	218 (10.9)	98 (11.4)	N/A	N/A	N/A
No	1782 (89.1)	765 (88.6)	N/A	N/A	N/A
Sexual orientation, n (%)	N/A	N/A	6.562 (1)	.01	0.048
Heterosexual	1805 (90.2)	751 (87.0)	N/A	N/A	N/A
Gay, bisexual, questioning, other	195 (9.8)	112 (13.0)	N/A	N/A	N/A
Household income, n (%)	N/A	N/A	31.505 (2)	<.001	0.105
Can $59,999 or less (US $46,153.07 or less)	747 (37.4)	238 (27.6)	N/A	N/A	N/A
Can $60,000 to $119,999 (US $46,153.86–92,306.92)	855 (42.8)	392 (45.4)	N/A	N/A	N/A
Can $120,000 or more (US $92,307.69 or more)	398 (19.9)	233 (27.0)	N/A	N/A	N/A

[a]DCM: Don't Change Much.

[b]As a ratio variable, the test performed was a *t* test for this characteristic.

[c]As a ratio variable, the test performed was a Cohen *d* for this characteristic.

[d]N/A: not applicable.

Compared with the benchmark no-exposure respondents, high-exposure respondents had significantly increased odds for all recent health behavior changes except *Quit or reduced smoking* (OR 0.820, 95% CI 0.461-1.458) and significantly decreased odds for *I haven't made any changes* (OR 0.140, 95% CI 0.065-0.301) while holding other predictor variables constant. Moderate effect sizes were observed for *Changed diet or improved eating habits* (OR 5.628, 95% CI 3.932-8.055), *Increased exercise, sports or physical activity* (OR 3.439, 95% CI 2.444-4.839), and *I haven't made any changes* (OR 0.140, 95% CI 0.065-0.301) [36]. Of the controlled predictor variables, age, employment, lives with children, education, and income were statistically significant with small effect sizes for some of the recent health behavior changes (see Table 2; full table in Multimedia Appendix 1).

Table 2. Logistic regressions between statistically significant demographics and recent health changes (separate multiple logistic regressions were conducted for each outcome variable with all predictor variables entered on the same step).

Dependent variables (recent health change)	Predictor variables, odds ratio (95% CI)							
	Don't Change Much level of exposure (ref=benchmark no exposure)			Age (years)	Employment	Education	Household income (ref=Can $60,000-$119,999; US $46,153.86–$92,306.92)	
	Limited exposure	Low exposure	High exposure				$59,999 or less (US $46,153.07 or less)	$120,000 or more (US $92,307.69 or more)
Changed diet or improved eating habits	2.008 (1.541-2.616)[a]	3.228 (2.588-4.027)[a]	5.628 (3.932-8.055)[a]	1.009 (1.002-1.015)[b]	1.136 (0.933-1.382)	0.792 (0.672-0.933)[b]	0.887 (0.734-1.074)	0.955 (0.776-1.175)
Made an effort to sit less and walk more	1.892 (1.442-2.481)[a]	3.079 (2.472-3.835)[a]	3.39 (2.457-4.678)[a]	1.018 (1.012-1.025)[a]	1.109 (0.902-1.363)	1.058 (0.893-1.253)	1.063 (0.87-1.299)	1.358 (1.101-1.676)[b]
Increased exercise, sports or physical activity	1.822 (1.398-2.375)[c]	2.064 (1.661-2.564)[c]	3.439 (2.444-4.839)[c]	0.992 (0.986-0.999)[a]	0.898 (0.743-1.086)	1.079 (0.921-1.265)	0.846 (0.703-1.018)	1.273 (1.042-1.557)[a]
I have not made any changes	0.39 (0.256-0.595)[c]	0.23 (0.154-0.344)[c]	0.14 (0.065-0.301)[c]	0.994 (0.986-1.002)	1.142 (0.893-1.461)	1.153 (0.936-1.421)	1.09 (0.857-1.385)	1.091 (0.834-1.426)
Drank less alcohol	1.486 (1.108-1.994)[b]	1.822 (1.439-2.306)[c]	3.287 (2.375-4.549)[c]	0.993 (0.986-1)	0.95 (0.765-1.179)	0.803 (0.668-0.964)[a]	1.031 (0.836-1.272)	0.876 (0.692-1.109)
Had a routine check-up or visit to doctor	1.855 (1.399-2.46)[c]	2.129 (1.693-2.677)[c]	2.082 (1.497-2.896)[c]	1.049 (1.041-1.056)[c]	0.572 (0.466-0.703)[c]	0.924 (0.777-1.1)	0.86 (0.7-1.056)	1.138 (0.919-1.41)
Improved consistent sleep quality	1.383 (1.014-1.885)[a]	1.781 (1.397-2.27)[c]	2.376 (1.696-3.329)[c]	1.001 (0.993-1.008)	1.014 (0.807-1.275)	0.994 (0.822-1.202)	1.113 (0.892-1.39)	1.051 (0.826-1.337)
Lost weight	1.255 (0.949-1.66)	1.602 (1.284-2.001)[c]	2.186 (1.59-3.005)[c]	1.002 (0.995-1.008)	1.106 (0.902-1.356)	0.84 (0.709-0.995)[a]	0.928 (0.761-1.131)	1.095 (0.887-1.352)
Reduced stress level	1.035 (0.758-1.415)	1.651 (1.305-2.089)[c]	1.945 (1.393-2.716)[c]	1.003 (0.996-1.01)	0.884 (0.712-1.097)	0.981 (0.818-1.177)	1.244 (1.007-1.537)[a]	0.981 (0.777-1.24)
Quit or reduced amount smoked	1.652 (1.162-2.347)[b]	0.944 (0.66-1.35)	0.82 (0.461-1.458)	0.988 (0.979-0.997)[b]	1.369 (1.025-1.827) [a]	0.447 (0.341-0.585)[c]	1.316 (1.006-1.721)[a]	0.6 (0.413-0.873)[b]

[a]$P<.05$.

[b]$P<.01$.

[c]$P<.001$.

Compared with the no-exposure respondents, high-exposure respondents had significantly increased odds for all intended health changes except *Quit or reduce amount smoked* (OR 1.043, 95% CI 0.559-1.946), while holding other predictor variables constant. Moderate effect sizes were observed for *Reduce stress level* (OR 4.282, 95% CI 3.086-5.941), *Improve consistent sleep quality* (OR 4.019, 95% CI 2.911-4.547), *Increase exercise, sports or physical activity* (OR 3.649, 95% CI 2.565-5.191), and *I don't intend to make any changes* (OR 0.162, 95% CI 0.082-0.321; see Table 3; full table in Multimedia Appendix 2).

A linear regression was, then, conducted. Compared with the no-exposure respondents, high-exposure respondents had significantly increased total number of recent and intended health changes, while holding other predictor variables constant. Moderate effect size (goodness of fit) was observed for both total number of recent health behavior changes ($F_{12,2850}$=25.52; $P<.001$; adjusted R^2=.093) and intended health behavior changes ($F_{12,2850}$=36.30; $P<.001$; adjusted R^2 =.129). The VIF criterion was within the acceptable range, and there was no indication of multicollinearity.

Table 3. Logistic regressions between statistically significant demographics and intended health changes (separate multiple logistic regressions were conducted for each outcome variable with all predictor variables entered on the same step).

Dependent variables (intended health change)	Predictor variables, odds ratio (95% CI)								
	Don't Change Much usage (ref=benchmark no exposure)			Age (years)	Employment	Education	Visible minority	Household income (ref=Can $60,000-$119,999; US $46,153.86–$92,306.92)	
	Limited exposure	Low exposure	High exposure					$59,999 or less (US $46,153.07 or less)	$120,000 or more (US $92,307.69 or more)
Improve consistent sleep quality	3.018 (2.3-3.959)[a]	3.633 (2.905-4.543) [a]	4.019 (2.911-5.547) [a]	0.992 (0.985-0.999)[b]	1.003 (0.81-1.242)	0.84 (0.703-1.004)	1.311 (1.008-1.706) [b]	0.961 (0.781-1.182)	1.019 (0.814-1.275)
Change diet or improve eating habits	3.671 (2.805-4.805)[a]	2.612 (2.097-3.254)[a]	3.244 (2.356-4.465) [a]	0.992 (0.985-0.998)[b]	0.94 (0.765-1.154)	0.826 (0.696-0.98)[b]	1.261 (0.976-1.629)	0.866 (0.709-1.058)	0.945 (0.762-1.172)
Increase exercise, sports or physical activity	2.707 (2.048-3.579) [a]	2.934 (2.336-3.687) [a]	3.649 (2.565-5.191) [a]	0.994 (0.988-1)[b]	1.029 (0.85-1.245)	0.928 (0.79-1.09)	1.295 (1.01-1.661)[b]	1.037 (0.86-1.249)	1.109 (0.904-1.361)
Make an effort to sit less and walk more	2.837 (2.168-3.713) [a]	3.253 (2.607-4.061) [a]	2.954 (2.144-4.069) [a]	1.013 (1.006-1.019) [a]	0.748 (0.61-0.918)[c]	0.816 (0.687-0.969)[b]	0.858 (0.652-1.128)	0.973 (0.796-1.189)	1.105 (0.891-1.371)
I don't intend to make any changes	0.223 (0.135-0.369)[a]	0.143 (0.089-0.23)[a]	0.162 (0.082-0.321)[a]	1.005 (0.997-1.013)	0.943 (0.743-1.195)	1.132 (0.921-1.392)	0.889 (0.642-1.231)	1.063 (0.839-1.347)	1.177 (0.904-1.534)
Reduce stress level	2.401 (1.803-3.197)[a]	2.337 (1.844-2.96)[a]	4.282 (3.086-5.941)[a]	0.983 (0.976-0.99)[a]	1.176 (0.934-1.482)	0.906 (0.751-1.093)	1.046 (0.79-1.384)	0.987 (0.792-1.229)	0.936 (0.738-1.189)
Lose weight	2.641 (2.016-3.458)[a]	2.4 (1.932-2.983)[a]	1.976 (1.44-2.713)[a]	1.005 (0.999-1.011)	1.057 (0.872-1.282)	0.8 (0.68-0.939)[c]	0.867 (0.674-1.114)	0.904 (0.75-1.091)	1.085 (0.886-1.328)
Have a routine check-up or visit to doctor	2.19 (1.624-2.953) [a]	2.024 (1.57-2.61) [a]	2.067 (1.443-2.961) [a]	1.021 (1.013-1.029) [a]	0.899 (0.71-1.139)	0.895 (0.734-1.092)	1.238 (0.913-1.678)	1.059 (0.842-1.332)	0.812 (0.627-1.05)
Drink less alcohol	1.677 (1.204-2.336)[c]	1.901 (1.453-2.489) [a]	2.974 (2.077-4.26) [a]	0.995 (0.987-1.003)	1.01 (0.782-1.304)	0.61 (0.491-0.757) [a]	1.228 (0.894-1.687)	0.731 (0.569-0.941)[b]	1.058 (0.813-1.376)
Quit or reduce smoking	1.415 (0.932-2.148)	0.829 (0.535-1.286)	1.043 (0.559-1.946)	0.987 (0.977-0.997)[b]	1.241 (0.896-1.72)	0.446 (0.324-0.614) [a]	0.664 (0.403-1.095)	1.236 (0.909-1.68)	0.504 (0.312-0.816)[b]

[a]*P*<.001.

[b]*P*<.05.

[c]*P*<.01.

Discussion

Principal Findings and Comparison With Previous Work

This study's findings confirm the potential of men's eHealth programs as previously described in a range of contexts and diverse studies [7,14-25]. Adding to the literature focused on understanding men's eHealth acceptability, engagement, and behavior change, this study contributes some important empirical insights and supports calls for future research to more fully investigate dose-response relationships with randomized controlled trials [25,26]. Although careful not to overstate the current findings or imply attribution, some explanations and potential implications for the statistically significant associations between men's DCM exposure levels and their recent and intended health behavior changes are offered as a means to scoping adjustments for DCM, and making broader recommendations for the men's eHealth field. Our hypothesis—increased DCM exposure levels would be

positively associated with men's recent and intended health behavior changes—was supported by the study results. These findings corroborate broader acceptability claims and previous reports [7,14-16] about the compatibility of diverse eHealth resources with some men's help-seeking preferences and practices. Also reinforced are assertions about the associations between engagement (based on self-reported duration and frequency of use) and men's recent and intended health behavior changes [22]. That these findings held when controlling for key demographics synonymous with social determinants of health (ie, income, employment, education) might be interpreted as reflecting the wide reach, accessibility, and engagement potential of DCM for men from diverse backgrounds. In essence, the DCM content and dissemination strategies seem to be acceptable to, and engaging of men from an array of circumstances, with the net result that the DCM components can support health behavior changes in wide-ranging end users. Future work might use structural equation modeling to investigate statistically significant predictor variables including age, education, employment, and household income to distil their mediation and moderating effects. That said, it is also important to acknowledge the possibility, within the context of this study, that motivation to make health behavior changes actually caused some men to engage with DCM.

In further breaking down this study's findings, questions emerge about if (and if so—how?) to act on some recent and intended health behavior change results to adjust the DCM content. Central here are the determinations for integrating specialist resources to DCM as a means to engaging more men with specific health behavior changes. For example, tobacco reduction and smoking cessation (TRSC) demand specialist resources (beyond the rhetoric messaging that *smoking is bad for you*) [15], and this study's findings regarding low recent and intended changes for *Quit or reduce smoking* likely reflect the relatively small number of male smokers in Canada (compared with the overall population), as well as the lack of dedicated DCM resources focused on men's TRSC. In essence, TRSC messaging was relevant to fewer end users (ie, smokers) and that relatively small subset of respondents were unlikely to have accessed DCM with the sole focus of reducing or quitting smoking. Similarly, that *Reduce stress* and *Improve consistent sleep quality* featured prominently as the most intended health behavior changes with fewer end users reporting recent changes in those health behaviors might indicate the need to adjust and/or integrate additional tailored DCM stress reduction and sleep aiding resources. Building on this point, Yardley et al [37] has argued the value of promoting effective engagement (defined empirically as sufficient engagement with the intervention to achieve intended outcomes) rather than simply more engagement. This is salient advice both in planning to adjust, add, and/or replace some DCM content and evaluating end users' experiences through triangulating time-based exposure data with qualitative interviews as a means to more fully contextualizing men's engagement. In addition, it is clear that longitudinal research is required to map content and end-user behavior changes over time and empirically guide ongoing adjustments to men's eHealth programs.

Although acknowledging the inherent complexities to accounting for human behaviors and health behavior change in men more specifically [38], this study's findings, consistent with findings from previous work [27,31], were mapped to the transtheoretical model (or stages of change). DCM respondents (contrasting the benchmark cohort) were clearly nestled across the contemplative, preparation, and action stages. That almost 10% and 13% of the variation in respondents' recent and intended health behaviour changes, respectively, were explained by DCM exposure levels and demographic variables, confirms the acceptability of, and engagement with DCM as well as the end users' readiness to change. Herein, DCM users can be broadly characterized as planning to make, as well as investing actions toward some health behavior changes. This finding confirms the DCM end users as a distinct subset of the male Canadian population, and although challenges remain for advancing more men past the precontemplative stage (toward DCM or similar), the DCM end users offer unique opportunities for building engagement, and by extension, aiding some men's efforts to maintain their health behavior changes.

Limitations

A methodological limitation suggesting caution for interpreting the results of this study is the high potential for familywise errors as a byproduct of conducting 22 separate regression analyses. Self-report biases, both in relation to respondents recalling their DCM usage and disclosing recent and intended health behavior changes are also limitations. In particular, social desirability especially pertaining to exercise and healthy eating may have influenced men's responses [39]. Reliance on quantitative measures limits the understanding about the diverse contexts that can influence men's health practices including their eHealth help-seeking. In addition, that respondents were Canadian reduces the generalizability of the findings to other men living elsewhere. Further acknowledged is that the benchmark sample was not stratified by race or ethnicity because defining stratification quotas by race/ethnicity in addition to gender, age, and location would have created too many interlocking stratification variables to administer and lead to sparse data within certain strata. This study, although purposefully differentiating acceptability and engagement, was also limited by its elementary conceptualization and formal evaluation of the acceptability of DCM [40]. Some of these limitations can, however, be addressed in future work by triangulating data collection to qualitatively build understandings about what constitutes and counts as engagement from end-user perspectives and more fully evaluating the multifaceted concept of acceptability [40], comparing men from other countries who visit DCM, and/or the use of larger sample sizes with noninterlocking strata for race/ethnicity. Although the hypothesis that increased DCM exposure levels were positively associated with men's recent and intended health behavior changes was supported by this study's findings, RCTs and control group comparisons are needed to make dose-response and attributions claims to advance the men's eHealth field. The inclusion of an economic analysis and specific behavior change techniques to consider cost would also strengthen future DCM studies [41].

Implications and Conclusions

Men's eHealth programs operate across a continuum of working to replace, augment, and connect men to professional *in person* health care services. This study, although focused on DCM in reporting men's recent and intended health behavior changes, confirms the potential of eHealth programs for aiding health behavior changes. Also offered are important empirical insights and approaches to designing content and evaluating men's eHealth resources, and mapping end-user outcomes with the transtheoretical model. Taken together, this study design and findings offer some methodological guidance and empirical weight to advance the men's eHealth field.

Acknowledgments

The Canadian Men's Health Foundation DCM eHealth program and DCM evaluation was supported with funding from the Public Health Agency of Canada. The DCM evaluation was subcontracted to, and completed by Intensions Consulting; The American Urologic Association, Research Scholar Award supported the involvement of RF.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Full table - Logistic regressions between demographics and recent health changes.
[DOCX File , 24 KB - jmir_v22i5e16174_app1.docx]

Multimedia Appendix 2
Full table - Logistic regressions between demographics and intended health changes.
[DOCX File , 24 KB - jmir_v22i5e16174_app2.docx]

References

1. Courtenay WH. Constructions of masculinity and their influence on men's well-being: a theory of gender and health. Soc Sci Med 2000 May;50(10):1385-1401. [doi: 10.1016/s0277-9536(99)00390-1] [Medline: 10741575]
2. O'Brien R, Hunt K, Hart G. 'It's caveman stuff, but that is to a certain extent how guys still operate': men's accounts of masculinity and help seeking. Soc Sci Med 2005 Aug;61(3):503-516. [doi: 10.1016/j.socscimed.2004.12.008] [Medline: 15899311]
3. Bilsker D, Goldenberg L, Davison J. Summit | SFU's Institutional Repository. 2010. A Roadmap to Men's Health: Current Status, Research, Policy & Practice URL: https://summit.sfu.ca/item/11205 [accessed 2020-02-10]
4. Oliffe JL, Robertson S, Frank B, McCreary DR, Tremblay G, Goldenberg SL. Men's health in Canada: a 2010 update. J Mens Health 2010;7(3):189-192. [doi: 10.1016/j.jomh.2010.07.001]
5. Robertson S, Baker P. Men and health promotion in the United Kingdom: 20 years further forward? Health Educ J 2017;76(1):102-113. [doi: 10.1177/0017896916645558]
6. Punjani N, Flannigan R, Oliffe JL, McCreary DR, Black N, Goldenberg SL. Unhealthy behaviors among Canadian men are predictors of comorbidities: Implications for clinical practice. Am J Mens Health 2018 Nov;12(6):2183-2193 [FREE Full text] [doi: 10.1177/1557988318799022] [Medline: 30222015]
7. Fogarty AS, Proudfoot J, Whittle EL, Clarke J, Player MJ, Christensen H, et al. Preliminary evaluation of a brief web and mobile phone intervention for men with depression: Men's positive coping strategies and associated depression, resilience, and work and social functioning. JMIR Ment Health 2017 Aug 10;4(3):e33 [FREE Full text] [doi: 10.2196/mental.7769] [Medline: 28798009]
8. Robinson M, Robertson S. Young men's health promotion and new information communication technologies: illuminating the issues and research agendas. Health Promot Int 2010 Sep;25(3):363-370. [doi: 10.1093/heapro/daq022] [Medline: 20427372]
9. Schnall R, Travers J, Rojas M, Carballo-Diéguez A. eHealth interventions for HIV prevention in high-risk men who have sex with men: a systematic review. J Med Internet Res 2014 May 26;16(5):e134 [FREE Full text] [doi: 10.2196/jmir.3393] [Medline: 24862459]
10. Canadian Men's Health Foundation. Don't Change Much. URL: https://www.dontchangemuch.ca/ [accessed 2020-02-10]
11. Lahoud R, Franco I. Cardiovascular prevention in men. In: Potts JM, editor. Men's Health: A Head to Toe Guide for Clinicians. New York: Springer; 2016:89-96.
12. Oliffe JL, McCreary DR, Black N, Flannigan R, Goldenberg SL. Canadian men's health literacy: a nationally representative study. Health Promot Pract 2019 Mar 19:1524839919837625. [doi: 10.1177/1524839919837625] [Medline: 30884981]
13. Oliffe JL, Rossnagel E, Bottorff JL, Chambers SK, Caperchione C, Rice SM. Community-based men's health promotion programs: eight lessons learnt and their caveats. Health Promot Int 2019 Oct 11:-. [doi: 10.1093/heapro/daz101] [Medline: 31603471]

14. Young MD, Morgan PJ. Effect of a gender-tailored eHealth weight loss program on the depressive symptoms of overweight and obese men: pre-post study. JMIR Ment Health 2018 Jan 9;5(1):e1 [FREE Full text] [doi: 10.2196/mental.8920] [Medline: 29317379]
15. Bottorff JL, Oliffe JL, Sarbit G, Sharp P, Caperchione CM, Currie LM, et al. Evaluation of QuitNow Men: An online, men-centered smoking cessation intervention. J Med Internet Res 2016 Apr 20;18(4):e83 [FREE Full text] [doi: 10.2196/jmir.5076] [Medline: 27097991]
16. Mackert M, Guadagno M, Lazard A, Donovan E, Rochlen A, Garcia A, et al. Engaging men in prenatal health via eHealth: findings from a national survey. JMIR Pediatr Parent 2018 Aug 9;1(2):e7 [FREE Full text] [doi: 10.2196/pediatrics.9513] [Medline: 31518311]
17. Da Costa D, Zelkowitz P, Letourneau N, Howlett A, Dennis C, Russell B, et al. HealthyDads.ca: what do men want in a website designed to promote emotional wellness and healthy behaviors during the transition to parenthood? J Med Internet Res 2017 Oct 11;19(10):e325 [FREE Full text] [doi: 10.2196/jmir.7415] [Medline: 29021126]
18. Ogrodniczuk JS, Oliffe JL, Beharry J. HeadsUpGuys: canadian online resource for men with depression. Can Fam Physician 2018 Feb;64(2):93-94 [FREE Full text] [Medline: 29449231]
19. Hirshfield S, Downing MJ, Parsons JT, Grov C, Gordon RJ, Houang ST, et al. Developing a video-based eHealth intervention for HIV-positive gay, bisexual, and other men who have sex with men: Study protocol for a randomized controlled trial. JMIR Res Protoc 2016 Jun 17;5(2):e125 [FREE Full text] [doi: 10.2196/resprot.5554] [Medline: 27315764]
20. Hirshfield S, Downing MJ, Chiasson MA, Yoon IS, Houang ST, Teran RA, et al. Evaluation of sex positive! A video eHealth intervention for men living with HIV. AIDS Behav 2019 Nov;23(11):3103-3118. [doi: 10.1007/s10461-019-02498-5] [Medline: 31011912]
21. van den Berg JJ, Silverman T, Fernandez MI, Henny KD, Gaul ZJ, Sutton MY, et al. Using eHealth to reach black and hispanic men who have sex with men regarding treatment as prevention and preexposure prophylaxis: Protocol for a Small Randomized Controlled Trial. JMIR Res Protoc 2018 Jul 16;7(7):e11047 [FREE Full text] [doi: 10.2196/11047] [Medline: 30012549]
22. Muessig KE, Bien CH, Wei C, Lo EJ, Yang M, Tucker JD, et al. A mixed-methods study on the acceptability of using eHealth for HIV prevention and sexual health care among men who have sex with men in China. J Med Internet Res 2015 Apr 21;17(4):e100 [FREE Full text] [doi: 10.2196/jmir.3370] [Medline: 25900881]
23. Bonett S, Connochie D, Golinkoff JM, Horvath KJ, Bauermeister JA. Paradata Analysis of an eHealth HIV testing intervention for young men who have sex with men. AIDS Educ Prev 2018 Oct;30(5):434-447 [FREE Full text] [doi: 10.1521/aeap.2018.30.5.434] [Medline: 30332307]
24. Klein CH, Kuhn T, Huxley D, Kennel J, Withers E, Lomonaco CG. Preliminary findings of a technology-delivered sexual health promotion program for black men who have sex with men: quasi-experimental outcome study. JMIR Public Health Surveill 2017 Oct 24;3(4):e78 [FREE Full text] [doi: 10.2196/publichealth.7933] [Medline: 29066422]
25. Forbes CC, Finlay A, McIntosh M, Siddiquee S, Short CE. A systematic review of the feasibility, acceptability, and efficacy of online supportive care interventions targeting men with a history of prostate cancer. J Cancer Surviv 2019 Feb;13(1):75-96 [FREE Full text] [doi: 10.1007/s11764-018-0729-1] [Medline: 30610736]
26. Simoni JM, Ronen K, Aunon FM. Health behavior theory to enhance eHealth intervention research in HIV: rationale and review. Curr HIV/AIDS Rep 2018 Dec;15(6):423-430 [FREE Full text] [doi: 10.1007/s11904-018-0418-8] [Medline: 30511186]
27. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. J Consult Clin Psychol 1983 Jun;51(3):390-395. [doi: 10.1037//0022-006x.51.3.390] [Medline: 6863699]
28. Prochaska JO, DiClemente CC, Norcross JC. In search of how people change. Applications to addictive behaviors. Am Psychol 1992 Sep;47(9):1102-1114. [doi: 10.1037//0003-066x.47.9.1102] [Medline: 1329589]
29. Pyky R, Koivumaa-Honkanen H, Leinonen A, Ahola R, Hirvonen N, Enwald H, et al. Effect of tailored, gamified, mobile physical activity intervention on life satisfaction and self-rated health in young adolescent men: A population-based, randomized controlled trial (MOPO study). Comput Hum Behav 2017 Jul;72:13-22. [doi: 10.1016/j.chb.2017.02.032]
30. Jeihooni AK, Rakhshani T. The effect of educational intervention based on health belief model and social support on promoting skin cancer preventive behaviors in a sample of Iranian farmers. J Cancer Educ 2019 Apr;34(2):392-401. [doi: 10.1007/s13187-017-1317-1] [Medline: 29313300]
31. Sanchez-Johnsen L, Craven M, Nava M, Alonso A, Dykema-Engblade A, Rademaker A, et al. Cultural variables underlying obesity in Latino men: Design, rationale and participant characteristics from the Latino Men's Health Initiative. J Community Health 2017 Aug;42(4):826-838 [FREE Full text] [doi: 10.1007/s10900-017-0324-9] [Medline: 28332066]
32. Flannigan RK, Oliffe JL, McCreary DR, Punjani N, Kasabwala K, Black N, et al. Composite health behaviour classifier as the basis for targeted interventions and global comparisons in men's health. Can Urol Assoc J 2019 Apr;13(4):125-132 [FREE Full text] [doi: 10.5489/cuaj.5454] [Medline: 30273120]
33. Quinn DM, Chaudoir SR. Living with a concealable stigmatized identity: the impact of anticipated stigma, centrality, salience, and cultural stigma on psychological distress and health. J Pers Soc Psychol 2009 Oct;97(4):634-651 [FREE Full text] [doi: 10.1037/a0015815] [Medline: 19785483]

34. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Second Edition. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
35. McHugh ML. The chi-square test of independence. Biochem Med (Zagreb) 2013;23(2):143-149 [FREE Full text] [doi: 10.11613/bm.2013.018] [Medline: 23894860]
36. Chen H, Cohen P, Chen S. How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. Commun Stat Simul Comput 2010 Mar 31;39(4):860-864. [doi: 10.1080/03610911003650383]
37. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
38. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring engagement in eHealth and mHealth behavior change interventions: Viewpoint of methodologies. J Med Internet Res 2018 Nov 16;20(11):e292 [FREE Full text] [doi: 10.2196/jmir.9397] [Medline: 30446482]
39. Couch D, Han G, Robinson P, Komesaroff P. Men's weight loss stories: how personal confession, responsibility and transformation work as social control. Health (London) 2019 Jan;23(1):76-96. [doi: 10.1177/1363459317724855] [Medline: 28803501]
40. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res 2017 Jan 26;17(1):88 [FREE Full text] [doi: 10.1186/s12913-017-2031-8] [Medline: 28126032]
41. Beard E, West R, Lorencatto F, Gardner B, Michie S, Owens L, et al. What do cost-effective health behaviour-change interventions contain? A comparison of six domains. PLoS One 2019;14(4):e0213983 [FREE Full text] [doi: 10.1371/journal.pone.0213983] [Medline: 30995224]

Abbreviations

DCM: Don’t Change Much
eHealth: electronic health
TRSC: tobacco reduction and smoking cessation
VIF: variance inflation factor

Edited by G Eysenbach; submitted 09.09.19; peer-reviewed by A Murphy, H McCall, S Hirshfield; comments to author 11.11.19; revised version received 03.01.20; accepted 28.01.20; published 15.05.20.

Please cite as:
Oliffe JL, Black N, Yiu J, Flannigan RK, McCreary DR, Goldenberg SL
Mapping Canadian Men’s Recent and Intended Health Behavior Changes Through the Don’t Change Much Electronic Health Program
J Med Internet Res 2020;22(5):e16174
URL: https://www.jmir.org/2020/5/e16174
doi:10.2196/16174
PMID:32412423

Original Paper

A Highly Tailored Text and Voice Messaging Intervention to Improve Medication Adherence in Patients With Either or Both Hypertension and Type 2 Diabetes in a UK Primary Care Setting: Feasibility Randomized Controlled Trial of Clinical Effectiveness

Aikaterini Kassavou[1], BSc, MSc, PhD; Venus Mirzaei[1], BSc, MSc; James Brimicombe[1], BA; Simon Edwards[1], BA; Efthalia Massou[1], BSc, MSc, PhD; A Toby Prevost[2], BSc, MSc, PhD; Simon Griffin[1], BSc, MBBS, MSc, MD; Stephen Sutton[1], BSc, MSc, PhD

[1]The University of Cambridge, Cambridge, United Kingdom
[2]Imperial College London, Cambridge, United Kingdom

Corresponding Author:
Aikaterini Kassavou, BSc, MSc, PhD
The University of Cambridge
Forvie Site
Robinson Way
Cambridge, CB2 0SR
United Kingdom
Phone: 44 1223330456
Email: kk532@medschl.cam.ac.uk

Abstract

Background: The efficacy of a highly tailored digital intervention to support medication adherence and feasibility to support clinical effectiveness as an adjunct to the primary care setting has not been evaluated.

Objective: This trial aimed to assess the behavioral efficacy of a highly tailored digital intervention to support medication adherence and to evaluate the feasibility of its clinical effectiveness, in patients with either or both hypertension and type 2 diabetes. We also examined quality of life and mechanisms of behavior change. Intervention fidelity, engagement, and satisfaction were also explored.

Methods: This was a multicenter, individually randomized controlled trial of 2 parallel groups: an intervention group that received a highly tailored text message and interactive voice response intervention for 12 weeks, and a control group that received usual care. Medication adherence was measured using self-reports and assessor-blinded practice records of a repeat prescription. Systolic blood pressure and glucose levels were assessed by nurses blinded to group allocation during practice visits at 3 months follow-up. Questionnaires obtained data to assess intervention mechanisms of action and satisfaction and digital log files captured data to evaluate fidelity and engagement.

Results: A total of 135 nonadherent patients (62/135, 46% female; 122/135, 90.3%; aged above 50 years) were randomly allocated in the intervention (n=79) or in the control group (n=56); of whom 13% (18/135) were lost at follow-up. Medication adherence was significantly improved in the intervention group compared with the control group (t_{116}=2.27; *P*=.02, 2-tailed). Systolic blood pressure was 0.6 mmHg (95% CI −7.423 to 6.301), and hemoglobin A_{1c} was 4.5 mmol/mol (95% CI −13.099 to 4.710) lower in the intervention group compared with the control group. Changes in intentional nonadherence and nonintentional nonadherence explained the improvements in medication adherence in the intervention group (beta=.074, SE=0.464; *P*=.04), but not in the control group (beta=.00, SE 1.35; *P*=.37). The intervention had 100% fidelity, a median of 12 days of engagement, and 76% overall satisfaction.

Conclusions: Our trial is the first that has been conducted in the United Kingdom and showed that among nonadherent patients with either or both hypertension and type 2 diabetes, a highly tailored digital intervention was effective at improving treatment adherence and feasible to obtain clinically meaningful outcomes. Changes in intentional and nonintentional nonadherence predicted the improvements in medication adherence. The intervention had high fidelity, engagement, and satisfaction. Future research using a rigorous design is needed to evaluate the clinical effectiveness and cost-effectiveness of the intervention in primary care.

Trial Registration: International Standard Randomized Controlled Trial Number (ISRCTN) 10668149; http://www.controlled-trials.com/ISRCTN10668149.

(J Med Internet Res 2020;22(5):e16629) doi:10.2196/16629

KEYWORDS

medication adherence; hypertension; type 2 diabetes; text messaging; interactive voice response

Introduction

Background

The clinical management of cardio-metabolic conditions, like hypertension and type 2 diabetes mellitus, is one of the most common consultations in primary care [1,2] and pharmacotherapy is an essential part of effective management of these conditions [3]. However, many patients do not adhere to their prescribed medication regimen [4], leading to reduced treatment efficacy, increased risk of complications, additional consultations, and hospital admissions [5,6]. Medication adherence can significantly reduce these risks and prevent health complications [7,8]. It could also result in health care savings, with estimates from the United Kingdom of approximately US $115.4 million preventable cost being spent annually, due to only nonadherence to antihypertensive medications [9].

Setting

Primary health care providers can facilitate treatment adherence, but their time is limited and expensive. Given the growing prevalence of hypertension, the associated comorbidities, and the aging population, it is likely that there will be an increased need for health care resources to support medication adherence. Digital interventions are promising strategies to support adherence [10,11] and can be highly tailored and acceptable adjunct to primary care consultations [12]. Telephone-based interventions that deliver text and voice messages may have a wider reach as they can be delivered to any device, even in places of low network coverage, and are accessible to people from all socioeconomic backgrounds and age groups. We have therefore developed the Medication Adherence for Patient's Support intervention, a behavioral intervention that uses text and voice messages, to support medication adherence in patients with either or both hypertension and type 2 diabetes as an adjunct to primary care consultations.

Aims and Objectives

This study aimed to assess if patients with either or both hypertension and type 2 diabetes, who used the intervention for 3 months in addition to usual care, had improved medication adherence and if they differed in terms of systolic blood pressure or hemoglobin A_{1c}, and quality of life at 3 months follow-up. We have also assessed intervention mechanisms of action, as well as intervention fidelity, engagement and satisfaction. However, this feasibility trial primarily aimed to attain evidence about the feasibility of obtaining medication-related clinical outcomes to inform a larger effectiveness and cost-effectiveness trial in primary care.

Methods

Study Design

We used an individually randomized controlled trial of unequal 3:2 allocation ratio with a 3-month follow-up. We used unequal allocation ratio to increase the information obtained about the intervention. Intervention group patients received highly tailored and interactive text and voice recognition messages for 12 weeks, as an adjunct to usual care. The control group received usual care only [13].

Study Setting

Participants were recruited over 7 months, from March to September 2018, from 8 primary care practices in the East of England. We selected a wide range of primary care practices from different areas of deprivation (based on Index of Multiple Deprivation), and at least 50% in highly deprived areas, to obtain information about the reach and implementation of the intervention in different sites, and to increase the scalability of the intervention to a larger trial. Follow-up data collection was conducted between May 2018 and February 2019.

Participant Recruitment and Eligibility Criteria

A member of the practice staff screened medical databases against the inclusion criteria to identify eligible patients. Patients were eligible when they met all the following criteria: (1) were 18 years or older, (2) had a diagnosis of hypertension, type 2 diabetes mellitus, or both health conditions; (3) had been prescribed at least 1 antihypertensive medication or glucose lowering medication as documented in practice records for a period of 3 to 6 months before recruitment; and (4) had either or both poorly controlled blood pressure (controlled by age group) and glucose levels as logged in their medical records, or had gaps in collecting repeat prescriptions during the 6 months before the study invitation. When possible, patients' digital literacy was screened for inclusion.

A general practitioner screened the generated list of potentially eligible patients against the exclusion criteria to confirm eligibility. Patients were excluded when they met any of the following criteria: (1) were taking part in another medication adherence or digital intervention or (2) had a health condition that could impair their participation in the study.

Eligible patients were invited to the study by the primary care practices using text messages, postal invitations, and follow-up calls to postage invitations, or were recruited opportunistically during usual care consultations (eg, blood pressure checks and medication reviews) by health care providers.

Study Procedures

Interested patients were invited and attended baseline consultations with a health care provider (a practice nurse or a health care facilitator), where they were screened for eligibility (eg, digital literacy) and asked to provide informed consent. Baseline consultations involved the completion of a baseline questionnaire, measuring medication adherence, quality of life, and theoretical determinants associated with intentional nonadherence and nonintentional nonadherence (eg, necessity beliefs about medication taking and self-efficacy) and were facilitated by the health care provider. Responses to the baseline questionnaire were included in a webpage (which consisted of the tailoring algorithm, the intervention schedule, the intervention messages, and the inbound calls) to further inform the tailored intervention to those allocated to receive the digital intervention.

After the end of the consultations, patients were randomized to intervention or control groups and were sent a text message or a letter with information about which group they are allocated in. Further, intervention group patients completed information (on a webpage or verbally during phone calls; depending on their digital literacy) about their preferred times and frequency to receive the intervention messages. They were also provided with more information on how to use the digital delivery modes. All patients were provided with additional information about the trial procedures. Baseline consultation was conducted few days before each patient's repeat prescription was due for collection.

Intervention

Intervention group patients received highly tailored and interactive text and voice recognition messages for 12 weeks, initiated the day their repeat prescription was due for collection. The intervention development was guided by the theoretical framework that distinguishes between intentional and nonintentional nonadherence [14], previous evidence [11,12,15], and included behavior change techniques and strategies mapped onto either or both intentional and nonintentional nonadherence. Nonintentional nonadherence refers to patients not taking their medications as prescribed because they forget or misunderstand the recommendations. Intentional nonadherence refers to patients not taking their medications as prescribed because they decide to take less medication or miss a dose or a day of their medications. The intervention aimed to improve medication adherence by employing behavior change strategies to modify either or both intentional and nonintentional nonadherence.

The interactive element of the intervention was utilized for dynamic tailoring: patients could record a personalized implementation intention plan to further tailor intervention content and be grouped to receive more habit formation or self-monitoring advice based on records and responses obtained during the intervention. We have also adopted a flexible approach to intervention delivery, providing participants with functions to change delivery options (eg, decrease or increase the frequency of messages or stop the messages). Details of the intervention development are published elsewhere [12].

Outcomes

This trial aimed to assess the behavioral efficacy of a highly tailored digital intervention to support medication adherence in patients with either or both hypertension and type 2 diabetes, and its feasibility to support medication-related clinical outcomes. We also examined mechanisms of action. Recruitment and attrition rate and intervention fidelity, engagement, and satisfaction were also reported.

Primary outcomes were the efficacy of the intervention to support medication adherence and the feasibility of its clinical effectiveness at a 3-month follow-up (T2). Secondary outcomes included the mechanisms by which the intervention supported medication adherence, as well as intervention engagement and satisfaction.

Efficacy was measured using 2 self-reported items of adherence (ie, days of adherence during the past week and percentage of adherence during the past month) at T2, and repeat prescription claims during the last 2 intervention months before T2. A subsample of patients (n=9 intervention and n=2 control group) were randomly selected and invited to use a Medication Event Monitoring System caps for a duration of 7 consecutive days after T2. Medication Event Monitoring System is an electronic pill-bottle cap that registers the date and time of bottle opening [16]. Patients were provided with information about how to use the Medication Event Monitoring System and the number and specific medications to include in the container during follow-up consultations (see Multimedia Appendix 1). Practice records raw data (ie, dates of prescription claims) were extracted by a member of the practice staff from medical records and Medication Event Monitoring System data were extracted by a member of the research team; both nursing and research team staff were blinded to group allocation.

Clinical outcomes were measured at T2 by nurses blinded to group allocation during practice visits. Blood pressure was measured in patients prescribed antihypertensive medications and hemoglobin A_{1c} from those prescribed glucose-lowering medications. Blood pressure was measured 3 times with 1-minute intervals using calibrated blood pressure digital monitor devices (eg, Omron). The last 2 readings were included in the analysis. Blood samples were anonymized and sent to the Addenbrookes Hospital Pathology laboratory for analysis of hemoglobin A_{1c}.

Quality of life was measured using the 5-level EQ-5D [17]. Medication Adherence Rating Scale [18] was used to measure intentional nonadherence (4-items) and nonintentional nonadherence (1-item). Feasibility was assessed by the recruitment and attrition rate (percentage of people who responded to the invitation and those randomized and the percentage of people who attended consultations at follow up).

Intervention fidelity was measured by calculating the proportion of messages received out of those messages scheduled per patient, for the duration of the 12 weeks intervention. Intervention engagement was calculated using the median number of days patients interacted with the intervention during the 12-week intervention period. Data were captured objectively

by digital log files and extracted by a member of the research team.

Sample Size

The sample was selected to obtain evidence about the feasibility to implement a cost-effectiveness trial in primary care. It aimed to provide evidence about the potential effectiveness of the intervention, and the values needed to estimate clinically meaningful outcomes (eg, mean and CIs for systolic blood pressure and hemoglobin A_{1c}). We included more participants in the intervention group, to obtain more information about the mechanisms by which the intervention supported medication adherence, and about the intervention engagement and satisfaction.

Randomization and Blinding

Block randomization (4 blocks, each of size of 25) was conducted to ensure random group allocation. The random sequence was generated by a centralized Web-based service [19] and was stratified in 2 important confounders: treatment adherence, as measured by the Medication Adherence Rating Scale, and burden of pills. Medication Adherence Rating Scale threshold of 24 was selected to indicate low (<24) or high (≥24) adherence. A burden of pills ratio of 10:6 (10 tablets: 6 different health conditions) was selected to indicate low (<10:6) or high (≥10:6) burden of pills; and the ratio was based on our pilot studies [12].

Data to calculate the burden of pills ratio were extracted from medical records, and Medication Adherence Rating Scale was self-reported. Both measures were entered into the randomization webpage by a member of the research team, after the completion of baseline consultations.

Health care providers were blinded to group allocation during baseline consultations. Participants were partly blinded during the baseline consultations as the purpose of the study was explained to them (trial videos in Multimedia Appendix 1).

Data Analysis

The analysis was performed based on complete cases at follow-up, excluding missing data. That is, only participants who completed the outcome measures at follow-up were included in the analysis. Histograms were used to explore continuous variables' distribution, and the Levene test was used to assess the assumption of equality of variance between groups. A *t*-test was used to investigate the differences between independent groups, and when its assumptions were not met, a nonparametric Mann-Whitney test for independent samples was conducted. Descriptive statistics (eg, means, SD, and percentages) was used for comparisons between groups, and for describing fidelity, engagement, and satisfaction with the intervention. A subgroup analysis was performed to explore differences in outcomes due to missing data: we compared if there were statistically significant differences between responders and nonresponders at each of the baseline and follow-up outcome measures. Subgroup analyses were performed on the total sample size. Multivariable regression analysis was performed to explore the mechanisms by which the intervention supported behavior change and control of medication-related clinical outcomes. The analysis was conducted in December 2019 using STATA.

Ethical Consideration

The study protocol was approved by the Ethics Committee of East of England, Essex Research Ethics Committee (REC Reference number 17/EE/0203) and Health Research Authority. Written informed consent was obtained from all study participants.

Results

Recruitment

A total of 4468 eligible patients were invited to the study. From those, 256 expressed interest to participate (a 5.7% response to the invitation), 140 were invited and attended consultations, and 135 were deemed eligible and randomized to the intervention (n=79) or control (n=56) group.

Randomization

The groups were very similar on baseline variables used for minimization and for most other variables. The majority (122/135, 90.3%) of patients were above 50 years of age (Table 1). Intervention group patients self-reported taking more pills per day compared with usual care group patients. This variable was not accounted for minimization, because an objective measure of burden of pills was selected (eg, burden of pills ratio extracted from medical records) instead of self-reports.

Table 1. Patients characteristics at baseline.

Variables[a]	Intervention group (n=79)	Usual care group (n=56)	*P* value
Index of multiple deprivation, n (%)			**N/A[b]**
20-30(most deprived)	9 (11.4)	6 (10.7)	
30-40	46 (58.3)	31 (55.3)	
40-50	0 (0)	0 (0)	
50-60	6 (7.6)	4 (7.2)	
60-70	6 (7.6)	8 (14.3)	
70-80 (least deprived)	12 (15.1)	7 (12.5)	
Age (years), n (%)			**.37**
18-29	0 (0)	1 (1.8)	
30-39	4 (5.1)	0 (0)	
40-49	5 (6.3)	3 (5.4)	
50-59	15 (19)	13 (23.2)	
60-69	27 (34.2)	15 (26.8)	
70-79	26 (32.9)	20 (35.7)	
80+	2 (2.5)	4 (7.1)	
Gender, n (%)			**.90**
Male	40 (50.7)	33 (59)	
Female	39 (49.3)	23 (41)	
Employment, n (%)			**.47**
Full-time	22 (27.8)	10 (17.8)	
Part-time	12 (15.2)	3 (5.4)	
Unemployed	1 (1.2)	1 (1.8)	
Unable to work due to disease	3 (3.8)	7 (12.5)	
Retired	41 (52)	35 (62.5)	
Number of pills prescribed to take per day, self-reported, mean (SD)	8.13 (6.21)	5.95 (5.14)	.03
Number of different pills prescribed to take per day, self-reported, mean (SD)	5.74 (3.67)	4.57 (3.52)	.06
Health condition, n (%)			**N/A**
Hypertension	42 (53.2)	38 (67.9)	
Type 2 diabetes	32 (40.5)	16 (28.6)	
Comorbidities of hypertension, type 2 diabetes, and cholesterol	5 (6.3)	2 (3.5)	
Medication adherence, mean (SD)			
Number of days of adherence, last week	6.46 (1.23)	6.38 (1.29)	.73
Percentage of adherence, last month	91.19 (13.11)	93.98 (33.32)	.49
Medication Adherence Rating Scale	22.78 (2.81)	23.25 (2.32)	.30
Repeat prescription	0.97 (0.20)	0.95 (0.26)	.64
Quality of life, mean (SD)			
EQ-5D[c], 5-items	1.71 (0.76)	1.74 (0.75)	.86
EQ-5D, total health	76.01 (19.68)	47.83 (21.56)	.74

[a]Data are reported as means (SD) or number (percentage). Health condition: main comorbidities were extracted from prescription data. Index of Multiple Deprivation was calculated based on general practice postcode; 10: low Index of Multiple Deprivation to 100: high Index of Multiple Deprivation. Repeat prescription was defined and calculated: supply of medication claimed by the patients, excluding the next prescription day the medication was claimed, and divided by the number of days of assessment period. Assessment period at baseline refers to the supply claimed by each patient before the

start of the study (ie, consent process). Assessment period at follow-up refers to the last 2 months of the study. Ratio was calculated per patient due to different denominators (eg, supply prescribed to be issued every 28 or 56 days). The overall adherence value was calculated by averaging each patient's ratio and dividing by the total number of patients.

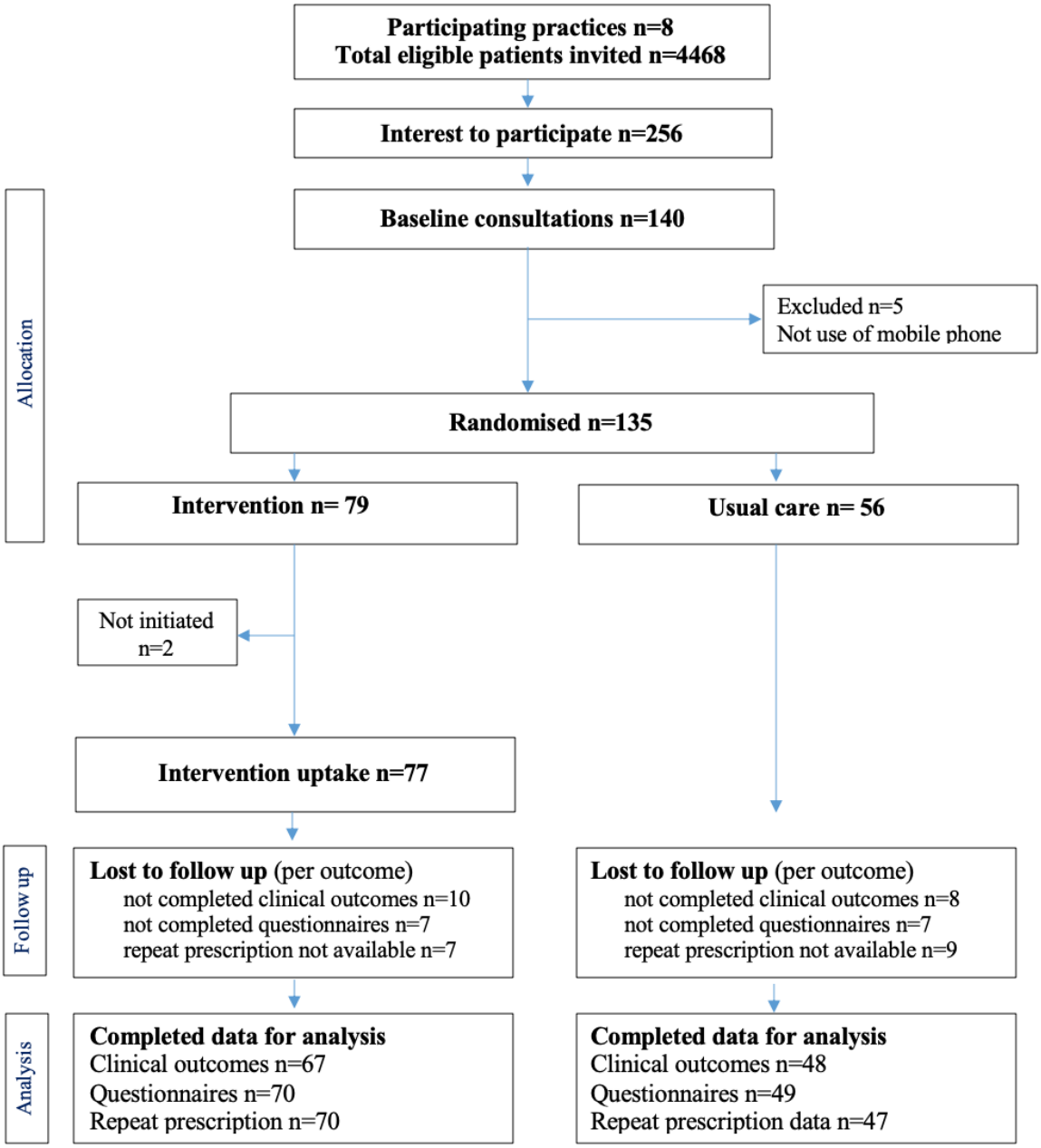
[b]N/A: not applicable.

[c]EQ-5D-5L: descriptive system of health-related quality of life states consisting of five dimension (mobility, self-care, usual activities, pain/discomfort, anxiety/depression) each of which can take one of five responses. The response record five levels of severity (no problems/slight problems/moderate problems/ severe problems/ extreme problems) within a particular EQ-5D dimension.

Attrition Rate

At 3 months follow-up, 88% (119/135) of participants completed the self-reported questionnaires and 85% (115/135) completed practice visits and provided clinical outcomes. Repeat prescription data were obtained for 87% (117/135) of the recruited patients (Figure 1).

Figure 1. Trial CONSORT patients flow diagram.



Intervention Retention Rate

Of those randomized to the intervention group, 97% (77/79) initiated the intervention: intervention uptake was defined by the number of participants who responded to the intervention messages during the first week of the intervention. Reported reasons for not initiating the intervention after group allocation was hospitalization. Intervention drop out was 2.5%, defined as the number of participants requesting to stop the intervention; and all drop out was captured during the first 4 weeks of the intervention.

Outcomes

The analysis found statistically significant difference in the repeat prescription adherence (t_{116}=2.27; *P*=.02; mean 0.99 [SD 0.11] for the intervention group vs 0.92 [SD 0.21] for the control group), days of adherence (t_{112}=2.37; *P*=.02; mean 6.85 [SD 0.47] vs 6.36 [SD 1.59]), Medication Event Monitoring System (t_{10}=4.04; *P*<.001; mean 6.05 [SD 2.29] vs 3.5 [SD 4.94]), and percentage of adherence (t_{112}=1.69; *P*=.05; mean 96.64 [SD 5.60] vs 91.89 [SD 18.60]) between groups, suggesting improvements in the intervention group. Nonstatistically significant differences in Medication Adherence Rating Scale (t_{104}=−0.24; *P*=.98; mean 23.66 [SD 1.99] vs 23.67 [SD 2.33]) between groups were found (Table 2).

Systolic blood pressure was lower (mean difference −0.6 mmHg, 95% CI −7.423 to 6.301) in the intervention group (137.8 mmHg) compared with the usual care group (138.4 mmHg). Similarly, hemoglobin A_{1c} was lower (a mean difference of −4.53 mmol/l, 95% CI −13.099 to 4.710) in the intervention group (57.2 mmol/l) in comparison to the usual care group (61.7 mmol/l), but for both of these clinical outcomes, this study was not intended to be powered to detect significant between-group differences. No statistically significant differences were found in quality of life between groups (t_{112}=0.524; *P*=.60; 2-tailed).

Subgroup analysis found no differences between the respondents and the nonrespondents of each of the above-mentioned outcomes.

Table 2. Difference between intervention and usual care group in medication adherence at 3 months follow-up.

Outcome[a]	Intervention group	Usual care group	Test statistic (*df*)	*P* value (2-tailed)	Mean difference	95% CI
Medication adherence						
Days adherence	6.85 (0.47)	6.36 (1.59)	2.37 (112)	.02	0.49	0.08 to 0.87
Percentage adherence	96.64 (5.60)	91.89 (18.59)	1.7 (112)	.05	4.74	−0.29 to 9.52
Repeat prescription	0.99 (0.11)	0.92 (0.21)	2.27 (116)	.02	0.07	0.00 to 0.12

[a]Figures per complete case analysis. Data are reported as means (SD) and are not adjusted for any baseline characteristics. The numbers present the difference between the intervention and the comparator group in outcomes at 3 months follow-up. Positive numbers present outcomes that were larger among the intervention group than the comparator group, and negative numbers present outcomes that were smaller among the intervention group than the comparator group.

Intervention Mechanism of Action

Multivariable regression analysis found that between-group differences in days of medication adherence was explained by changes in intentional and nonintentional nonadherence for the intervention group (beta.=074, SE=0.464; *P*=.04), but not for the control group (beta=.00, SE=1.35; *P*=.37); suggesting that changes in intentional and nonintentional nonadherence explained the improvements in medication adherence for the intervention group.

Further multivariable regression analysis revealed that better control of clinical outcomes (blood pressure <140/90mmHg or hemoglobin A_{1c} <42mmol/mol) was predicted by more days of medication adherence and positive beliefs about taking medication for the intervention group (beta=3.57, *P*=.03), but not for the control group (beta=.795, *P*=.46).

Intervention Fidelity, Engagement, and Satisfaction

Intervention fidelity was 100% for the total duration of the intervention, suggesting that all messages were delivered as scheduled. Most (56/77, 73%) intervention group patients selected to receive 1 intervention message per day and some (21/77, 27%) selected to receive 2 messages per day. The majority of patients selected to receive intervention messages around the time they used to take their medications.

Total intervention engagement was 12 days, suggesting that patients interacted with the digital intervention and potentially engaged with the medication adherence intervention.

The majority of the participants found the intervention easy to use (51/70, 73%), liked the automated voice delivering the voice messages (44/70, 62.5%), the content of the messages (50/70, 71.5%), and the availability to call and ask questions when needed (36/70, 51.6%). Overall, patients were satisfied with the experience with the intervention (53/70, 76%), and they would recommend it to other people who take medications for a long-term health condition (46/70, 65%; Multimedia Appendix 2).

Discussion

Principal Findings

Among those patients with either or both high blood pressure and high glucose levels recruited from primary care practices, patients who used the digital intervention improved their adherence to medication at 3 months by an average of 2 days (intervention effect size Cohen d=0.42), compared with those continuing with their usual care only. Taking into consideration that adherence declines during the first month following usual care consultations [20], our trial suggests that the intervention can support treatment adherence to nonadherent patient.

The attrition rate was low, and the end of intervention clinical outcomes found that the intervention group patients had lower blood pressure and lower glucose levels, compared with those allocated in the control group. Although the study was not powered to assess effectiveness in clinical outcomes, considering the impact of medication nonadherence on uncontrolled blood pressure [21] and uncontrolled hemoglobin A_{1c} [22,23], these results provide us with confidence that the intervention is feasible and could potentially be an effective adjunct to primary care consultation.

The intervention effectively improved medication adherence by modifying patients' intentional and nonintentional nonadherence, which proves the importance of the highly tailored intervention to support health behavior change. This result proves that this highly tailored intervention can effectively support the processes of behavioral change and in turn the health behavioral outcomes.

It was also found that patients interacted with the digital intervention for 12 days, which provides us with confidence that patients engaged with the medication adherence intervention [24], a prerequisite to achieve behavior change and clinically meaningful outcomes at least in the short term.

Strengths and Limitations

This trial provides evidence about the efficacy of the intervention to improve medication adherence based on self-reports, which might over estimate adherence levels. Nevertheless, the proxy measure of adherence, that is, the repeat prescription claims, was statistically significant, which provides us with confidence about the efficacy of the intervention; although both these measures do not directly reflect medication-taking behavior.

Another strength of the study is the recruitment of patients from a range of primary care practices, most in highly deprived areas. Considering the challenges with recruitment in a primary care setting [25], the recruitment rate provided us with confidence that the interventions can be scaled up to a wider range of practices and reach patients of different socioeconomic backgrounds.

A limitation of this study is that we could not reliably identify the patterns of nonadherent behavior (eg, frequency and sequence of nonadherent behavior) when patients were enrolled in the study. Similarly, we could not reliably identify the frequency of changes in prescribed medications, as well as the reasons for these changes (eg, therapeutic thresholds reported by patients or clinicians) [26]. Nevertheless, we decided to enroll all patients, regardless of the patterns or reasons of past nonadherence, to obtain more information about the intervention engagement and mechanism of action.

Possible Mechanisms and Implications for Clinicians or Policymakers

The results of this study suggest that the intervention was effective at improving medication adherence by supporting intentional and nonintentional nonadherence, had good engagement and satisfaction, and low attrition rate. In view of the demographics of the participants, who had more experience with and possibly expectation of patient-clinician interactions, and the recruitment setting of the study [25], these findings prove that the digital intervention is a feasible and acceptable adjunct to primary care consultation.

The baseline face-to-face consultation lasted on average 30 min and involved the completion of study procedures and signposting patients to the digital intervention, in addition to usual care (eg, addressing patients' concerns and side effects). From those patients allocated to the digital intervention, only 1% requested additional human support regarding the use of this technology, and requests involved changes to the time of the message delivery, or the dose or the name of the medication when a change in the prescribed tablet was recommended by their health care providers. Moreover, the low attrition rate suggests that the intervention is highly acceptable as an adjunct to usual care. Given that the cost to deliver the digital intervention is very little (monthly cost of £0.029 per patient), and there were no adverse events or additional consultations recorded during the 3 months, it is likely that the intervention is inexpensive for the primary care.

Future Directions

Improved medication adherence has been associated with decrease in all-cause mortality and morbidity, and digital interventions can be a solution to rapidly evolve the provision of health care [10]. Future research of larger sample sizes and objective outcome measures for medication adherence (eg, urine analysis for detection of antihypertensive medication) is needed to provide evidence about the effectiveness and cost-effectiveness of the intervention to support treatment adherence and clinical outcomes.

Conclusions

This is the first trial that has been conducted in the primary care setting in the United Kingdom. The finding of this feasibility trial suggests that the intervention is a feasible adjunct to primary care consultations and effective in improving medication adherence and process of care to patients nonadherent to either or both antihypertensive and antiglycaemic medications. The intervention is scalable and low cost. More research is needed to evaluate its effectiveness and cost effectiveness.

Acknowledgments

This paper presents independent research funded by the National Institute for Health Research (NIHR) under the Research for Patient Benefit program [grant number PB-PG-0215-36032]. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. Authors AK, VM, JB, SE, SG, TP, and SS were funded by the grant PB-PG-0215-36032 and EM from NIHR Research Capability Funding. No financial disclosures were reported by the authors of this paper.

The authors would like to acknowledge Vikki Houghton for assisting in data collection. We would also like to thank the following GP practices in East of England for taking part in this trial: the Breckland Alliance (Grove Surgery, Watton Medical Practice and School Lane Surgery), the Cornerstone Practice, the Little St Jones Street Surgery, the Mattishal Surgery, the New Queen Street Surgery and the Stanground Surgery, the Orchard Surgery Dereham, the Riverside Practice, and the Wansford Surgery.

Authors' Contributions

AK and SS have developed the intervention. JB and SE developed the digital platforms to facilitate intervention delivery in collaboration with AK. SG provided clinical advice and TP statistical advice during the design of this trial. EM provided statistical advice during data analysis. AK led the implementation and evaluation of this trial in the primary care, and VM assisted in data collection. AK conducted the analysis and drafted this publication. SS advised AK during the above mentioned processes. The paper's contents have not been previously presented elsewhere. All authors have read and approved this manuscript for publication and report no financial disclosures. Primary data are available upon request to AK.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Trial videos.
[DOCX File , 12 KB - jmir_v22i5e16629_app1.docx]

Multimedia Appendix 2
Usability, satisfaction and mechanisms of action.
[DOCX File , 16 KB - jmir_v22i5e16629_app2.docx]

Multimedia Appendix 3
CONSORT-eHEALTH (V 1.6.1).
[PDF File (Adobe PDF File), 381 KB - jmir_v22i5e16629_app3.pdf]

References

1. NHS Digital.: Health & Social Care Information Centre; 2018 Oct 26. Quality and Outcomes Framework (QOF), England: 2017-18 Prevalence, Achievement and Exceptions Report URL: https://files.digital.nhs.uk/B9/294E0E/qof-1718-rep.pdf [accessed 2018-11-10]
2. Cassell A, Edwards D, Harshfield A, Rhodes K, Brimicombe J, Payne R, et al. The epidemiology of multimorbidity in primary care: a retrospective cohort study. Br J Gen Pract 2018 Apr;68(669):e245-e251 [FREE Full text] [doi: 10.3399/bjgp18X695465] [Medline: 29530918]
3. National Institute of Health and Care Excellence. 2016 Nov. Hypertension in Adults: Diagnosis and Management URL: https://www.nice.org.uk/guidance/cg127 [accessed 2017-01-23]
4. Mazzaglia G, Ambrosioni E, Alacqua M, Filippi A, Sessa E, Immordino V, et al. Adherence to antihypertensive medications and cardiovascular morbidity among newly diagnosed hypertensive patients. Circulation 2009 Oct 20;120(16):1598-1605. [doi: 10.1161/CIRCULATIONAHA.108.830299] [Medline: 19805653]
5. Breekveldt-Postma NS, Beest FJ, Siiskonen SJ, Falvey H, Vincze G, Klungel OH, et al. The effect of discontinuation of antihypertensives on the risk of acute myocardial infarction and stroke. Curr Med Res Opin 2008 Jan;24(1):121-127. [doi: 10.1185/030079908x253843] [Medline: 18031596]
6. Kettani F, Dragomir A, Côté R, Roy L, Bérard A, Blais L, et al. Impact of a better adherence to antihypertensive agents on cerebrovascular disease for primary prevention. Stroke 2009 Jan;40(1):213-220. [doi: 10.1161/STROKEAHA.108.522193] [Medline: 19038916]
7. Chowdhury R, Khan H, Heydon E, Shroufi A, Fahimi S, Moore C, et al. Adherence to cardiovascular therapy: a meta-analysis of prevalence and clinical consequences. Eur Heart J 2013 Oct;34(38):2940-2948. [doi: 10.1093/eurheartj/eht295] [Medline: 23907142]
8. Esposti LD, Saragoni S, Benemei S, Batacchi P, Geppetti P, Di Bari M, et al. Adherence to antihypertensive medications and health outcomes among newly treated hypertensive patients. Clinicoecon Outcomes Res 2011;3:47-54 [FREE Full text] [doi: 10.2147/CEOR.S15619] [Medline: 21935332]
9. York Health Economics Consortium, School of Pharmacy, University of London. Evaluation of the Scale, Causes and Costs of Waste Medicines. London: Department of Health; 2010.
10. WHO. WHO Guidelines, Recommendations on Digital Interventions for Health System Strengthening. Geneva: World Health Organisation; 2019.
11. Kassavou A, Sutton S. Automated telecommunication interventions to promote adherence to cardio-metabolic medications: meta-analysis of effectiveness and meta-regression of behaviour change techniques. Health Psychol Rev 2018 Mar;12(1):25-42. [doi: 10.1080/17437199.2017.1365617] [Medline: 28805162]
12. Kassavou A, Houghton V, Edwards S, Brimicombe J, Sutton S. Development and piloting of a highly tailored digital intervention to support adherence to antihypertensive medications as an adjunct to primary care consultations. BMJ Open 2019 Jan 6;9(1):e024121 [FREE Full text] [doi: 10.1136/bmjopen-2018-024121] [Medline: 30613027]

13. The National Institute for Health and Care Excellence (NICE). 2009. Medicines Adherence: Involving Patients in Decisions About Prescribed Medicines and Supporting Adherence URL: https://www.nice.org.uk/guidance/cg76 [accessed 2020-03-23]
14. Barber N. Should we consider non-compliance a medical error? Qual Saf Health Care 2002 Mar;11(1):81-84 [FREE Full text] [doi: 10.1136/qhc.11.1.81] [Medline: 12078377]
15. Kassavou A, Sutton S. Reasons for non-adherence to cardiometabolic medications, and acceptability of an interactive voice response intervention in patients with hypertension and type 2 diabetes in primary care: a qualitative study. BMJ Open 2017 Aug 11;7(8):e015597 [FREE Full text] [doi: 10.1136/bmjopen-2016-015597] [Medline: 28801402]
16. Sutton S, Kinmonth A, Hardeman W, Hughes D, Boase S, Prevost AT, et al. Does electronic monitoring influence adherence to medication? Randomized controlled trial of measurement reactivity. Ann Behav Med 2014 Dec;48(3):293-299 [FREE Full text] [doi: 10.1007/s12160-014-9595-x] [Medline: 24573909]
17. EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. Health Policy 1990 Dec;16(3):199-208. [doi: 10.1016/0168-8510(90)90421-9] [Medline: 10109801]
18. Mora PA, Berkowitz A, Contrada RJ, Wisnivesky J, Horne R, Leventhal H, et al. Factor structure and longitudinal invariance of the Medical Adherence Report Scale-Asthma. Psychol Health 2011 Jun;26(6):713-727. [doi: 10.1080/08870446.2010.490585] [Medline: 21391132]
19. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. Biometrics 1975 Mar;31(1):103-115. [doi: 10.2307/2529712] [Medline: 1100130]
20. Naderi SH, Bestwick JP, Wald DS. Adherence to drugs that prevent cardiovascular disease: meta-analysis on 376,162 patients. Am J Med 2012 Sep;125(9):882-7.e1. [doi: 10.1016/j.amjmed.2011.12.013] [Medline: 22748400]
21. Bramley TJ, Gerbino PP, Nightengale BS, Frech-Tamas F. Relationship of blood pressure control to adherence with antihypertensive monotherapy in 13 managed care organizations. J Manag Care Pharm 2006 Apr;12(3):239-245 [FREE Full text] [doi: 10.18553/jmcp.2006.12.3.239] [Medline: 16623608]
22. Farmer AJ, Rodgers LR, Lonergan M, Shields B, Weedon MN, Donnelly L, et al. Adherence to oral glucose-lowering therapies and associations with 1-year HbA1c: a retrospective cohort analysis in a large primary care database. Diabetes Care 2016 Feb;39(2):258-263 [FREE Full text] [doi: 10.2337/dc15-1194] [Medline: 26681714]
23. Pladevall M, Williams LK, Potts LA, Divine G, Xi H, Lafata JE. Clinical outcomes and adherence to medications measured by claims data in patients with diabetes. Diabetes Care 2004 Dec;27(12):2800-2805 [FREE Full text] [doi: 10.2337/diacare.27.12.2800] [Medline: 15562188]
24. Kassavou A, Turner A, Hamborg T, French DP. Predicting maintenance of attendance at walking groups: testing constructs from three leading maintenance theories. Health Psychol 2014 Jul;33(7):752-756. [doi: 10.1037/hea0000015] [Medline: 24245833]
25. Macharia W, Leon G, Rowe B, Stephenson BJ, Haynes RB. An overview of interventions to improve compliance with appointment keeping for medical services. J Am Med Assoc 1992 Apr 1;267(13):1813-1817. [Medline: 1532036]
26. Burnier M. Is there a threshold for medication adherence? Lessons learnt from electronic monitoring of drug adherence. Front Pharmacol 2018;9:1540 [FREE Full text] [doi: 10.3389/fphar.2018.01540] [Medline: 30687099]

Abbreviations

NIHR: National Institute for Health Research

Edited by E Borycki; submitted 09.10.19; peer-reviewed by B Smith, E Rutebemberwa; comments to author 20.12.19; revised version received 04.01.20; accepted 01.02.20; published 19.05.20.

Please cite as:
Kassavou A, Mirzaei V, Brimicombe J, Edwards S, Massou E, Prevost AT, Griffin S, Sutton S
A Highly Tailored Text and Voice Messaging Intervention to Improve Medication Adherence in Patients With Either or Both Hypertension and Type 2 Diabetes in a UK Primary Care Setting: Feasibility Randomized Controlled Trial of Clinical Effectiveness
J Med Internet Res 2020;22(5):e16629
URL: http://www.jmir.org/2020/5/e16629/
doi:10.2196/16629
PMID:32427113

Original Paper

User Experiences of an Internet-Based Stepped-Care Intervention for Individuals With Cancer and Concurrent Symptoms of Anxiety or Depression (the U-CARE AdultCan Trial): Qualitative Study

Helena Igelström[1,2], PhD; Anna Hauffman[2], MSc; Sven Alfonsson[3,4], PhD; Jonas Sjöström[5], PhD; Åsa Cajander[6], PhD; Birgitta Johansson[2], PhD

[1]Department of Neuroscience, Uppsala University, Uppsala, Sweden
[2]Department of Immmunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden
[3]Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden
[4]Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden
[5]Department of Informatics and Media, Uppsala University, Uppsala, Sweden
[6]Department of Information Technology, Uppsala University, Uppsala, Sweden

Corresponding Author:
Helena Igelström, PhD
Department of Neuroscience
Uppsala University
BMC, Box 593
Uppsala, 75124
Sweden
Phone: 46 184714761
Email: helena.igelstrom@neuro.uu.se

Abstract

Background: The internet-based stepped-care intervention iCAN-DO, used in the multicenter randomized controlled trial *AdultCan*, was developed for adult patients undergoing treatment for cancer and concurrently experiencing anxiety or depressive symptoms. iCAN-DO aimed to decrease symptoms of anxiety or depression. Step 1 comprises access to a library with psychoeducational material and a peer-support section, as well as the possibility to pose questions to a nurse. Step 2 of the intervention offers treatment consisting of internet-based cognitive behavioral therapy (iCBT) to participants still experiencing anxiety or depression at 1, 4, or 7 months after inclusion.

Objective: The study aimed to explore user experiences of delivery, design, and structure of iCAN-DO from the perspective of people with cancer.

Methods: We studied user experiences by interviewing 15 informants individually: 10 women with breast cancer (67%), 4 men with prostate cancer (27%), and 1 man with colorectal cancer (7%) with a mean age 58.9 years (SD 8.9). The interviews focused on informants' perceptions of ease of use and of system design and structure. Informants had been included in iCAN-DO for at least 7 months. They were purposefully selected based on activity in Step 1, participation in iCBT (ie, Step 2), gender, and diagnosis.

Results: Of the 15 informants, 6 had been offered iCBT (40%). All informants used the internet on a daily basis, but 2 (13%) described themselves as very inexperienced computer users. The analysis revealed three subthemes, concerning how user experiences were affected by disease-specific factors and side effects (User experience in the context of cancer), technical problems (Technical struggles require patience and troubleshooting), and the structure and design of iCAN-DO (Appealing and usable, but rather simple).

Conclusions: The results indicate that user experiences were affected by informants' life situations, the technical aspects and the design of iCAN-DO, and informants' preferences. The results have generated some developments feasible to launch during the ongoing study, but if iCAN-DO is to be used beyond research interest, a greater level of tailoring of information, features, and design may be needed to improve user experiences. The use of recurrent questionnaires during the treatment period may highlight an individual's health, but also function as a motivator showing improvements over time.

(J Med Internet Res 2020;22(5):e16604) doi:10.2196/16604

KEYWORDS

interactive web portal; stepped care; user experience; cancer; interviews

Introduction

Today, many people with cancer are cared for as outpatients during cancer treatment [1], leaving limited occasions for meeting health care personnel. People with cancer may experience multiple symptoms from the disease and treatment [2] and might need information and advice for self-care of these symptoms [3]. Many patients have questions, especially during ongoing treatment, and would need information and advice, but there is a risk that these questions remain unanswered and patients live with unmet needs [4-6].

One option is to have patients access information and support functions via the internet. Internet-based support can provide relevant quality-assured information with just-in-time access, and can be combined with packages for enhanced social support and behavior change [7]. One advantage with this is that people can use such support when they feel ready for it or have time to access a computer or mobile device.

Internet-based support for people with cancer could be a way of providing information and support that may improve self-management of symptoms and thereby relieve distress and unmet needs. Further, it seems that improved access to information via the internet can stimulate engagement and empowerment [8]. In *AdultCan* [9], which is a multicenter randomized controlled trial (RCT), an internet-based stepped-care intervention, iCAN-DO, has been evaluated regarding its effects on anxiety and depressive symptoms in people with breast, colorectal, or prostate cancer with these symptoms.

If iCAN-DO is effective in reducing symptoms, the step toward clinical implementation is not far ahead. However, before implementing internet-based support into clinical care, the experience must be carefully explored from the user perspective. Important user-experience aspects to explore may be ease of use, navigation in the system, and perceptions of the design and structure [10].

This paper aims at exploring user experiences of delivery, design, and structure of the internet-based stepped-care intervention iCAN-DO used in *AdultCan* from the perspective of individuals with cancer who have symptoms of anxiety or depression.

Methods

Design

This was a qualitative study with an inductive approach using data collected through semistructured individual interviews.

Setting

The Uppsala University Psychosocial Care Program and the Web Portal

We recruited the informants in this study from *AdultCan* (ClinicalTrials.gov Identifier: NCT-01630681), an RCT within the Uppsala University Psychosocial Care Program (U-CARE) [11]. U-CARE is a strategic research venture supported by the Swedish government and has a multidisciplinary composition involving caring science, psychology, and information systems. All U-CARE research projects deliver interventions and collect patient-reported data via a web portal (ie, *the portal*) developed within the U-CARE program.

At the time of the start of *AdultCan* in 2013, end users could access the portal using a two-step verification: (1) via username and password and (2) with a code sent by SMS. Most of the participant views employed the Bootstrap framework.

The AdultCan Trial

AdultCan targeted individuals with breast, colorectal, or prostate cancer with symptoms of anxiety or depression. Screening regarding symptoms of anxiety and depression was performed using the Hospital Anxiety and Depression Scale (HADS) [12]; participants with a score of more than 7 on either of the two subscales were randomized to the iCAN-DO or standard care groups.

We developed the structure and content of iCAN-DO in collaboration with team members in clinical cancer care, system developers, and end users [13]. Step 1 of the intervention was accessible for 24 months and comprised a nurse-led interactive support program based on Orem's Self-Care Deficit Theory and the concept of psychoeducation [14] and social cognitive theory [15]. A key part in this support program was an information library with materials and self-care advice concerning common problems surrounding cancer, such as pain, nausea, and sleeping problems (see Figures 1-4). Figure 1 illustrates the start page in the portal while the subsequent figures illustrate visual and written information on diagnosis (see Figure 2), treatment (see Figure 3), and advice for symptom management (see Figure 4).

Figure 1. The start page of the portal (in Swedish).

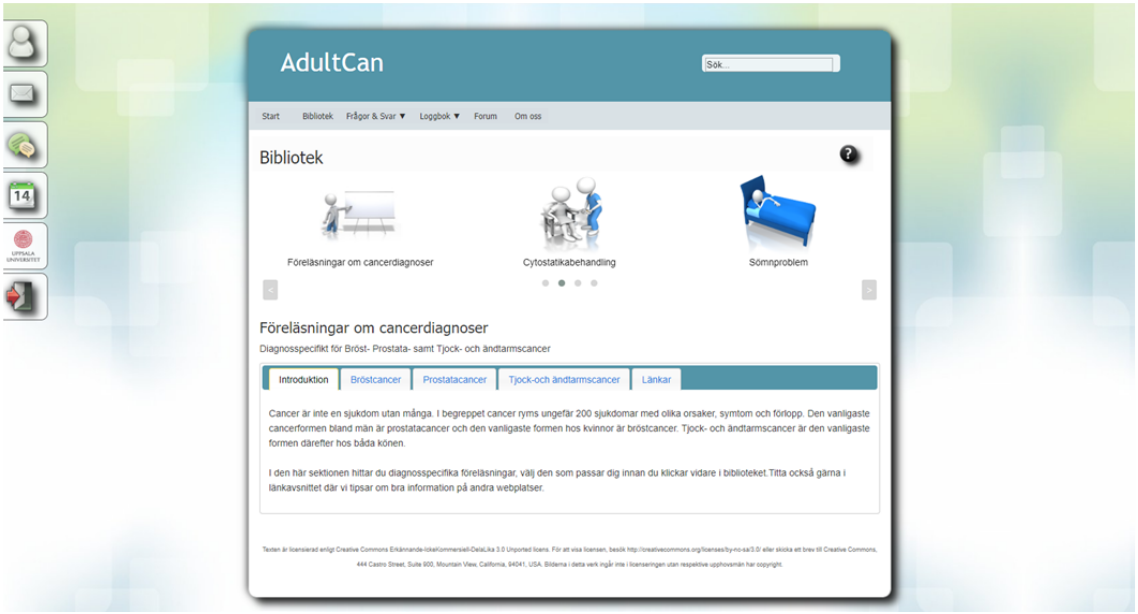
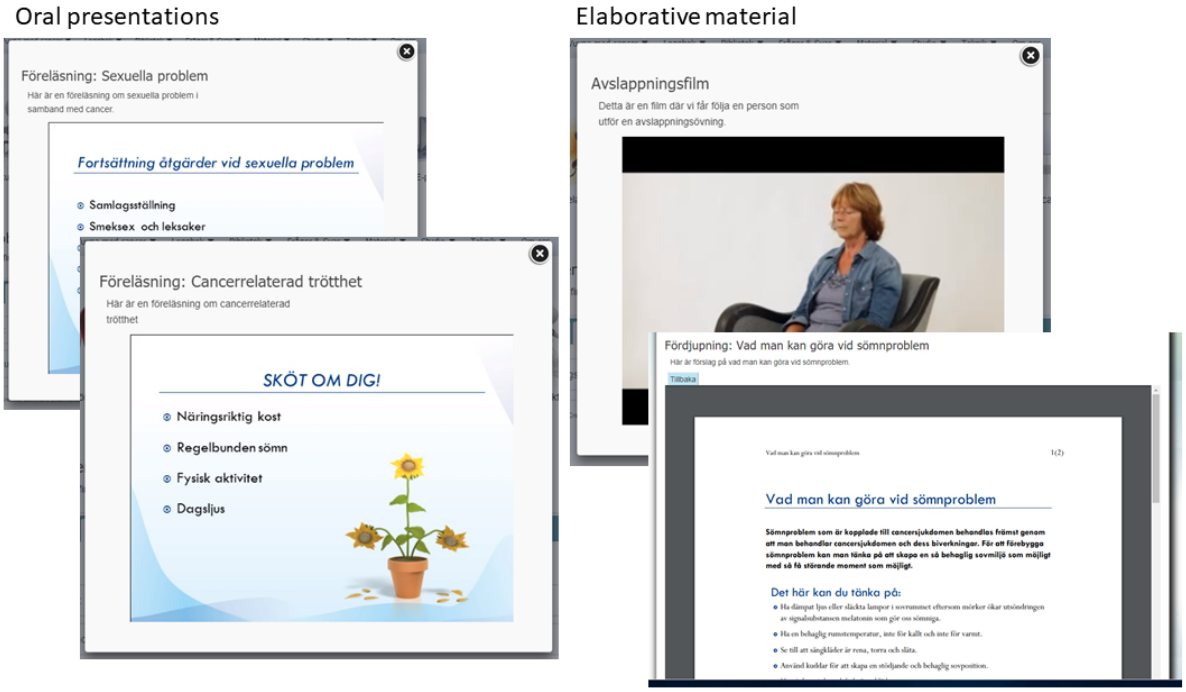


Figure 2. Diagnoses, with information on pathophysiology, prevalence, incidence, and prognosis (in Swedish).

Figure 3. Oncological treatments: why and how (in Swedish).

Figure 4. Symptoms and consequences of anxiety and depression as well as advice for self-care (in Swedish).

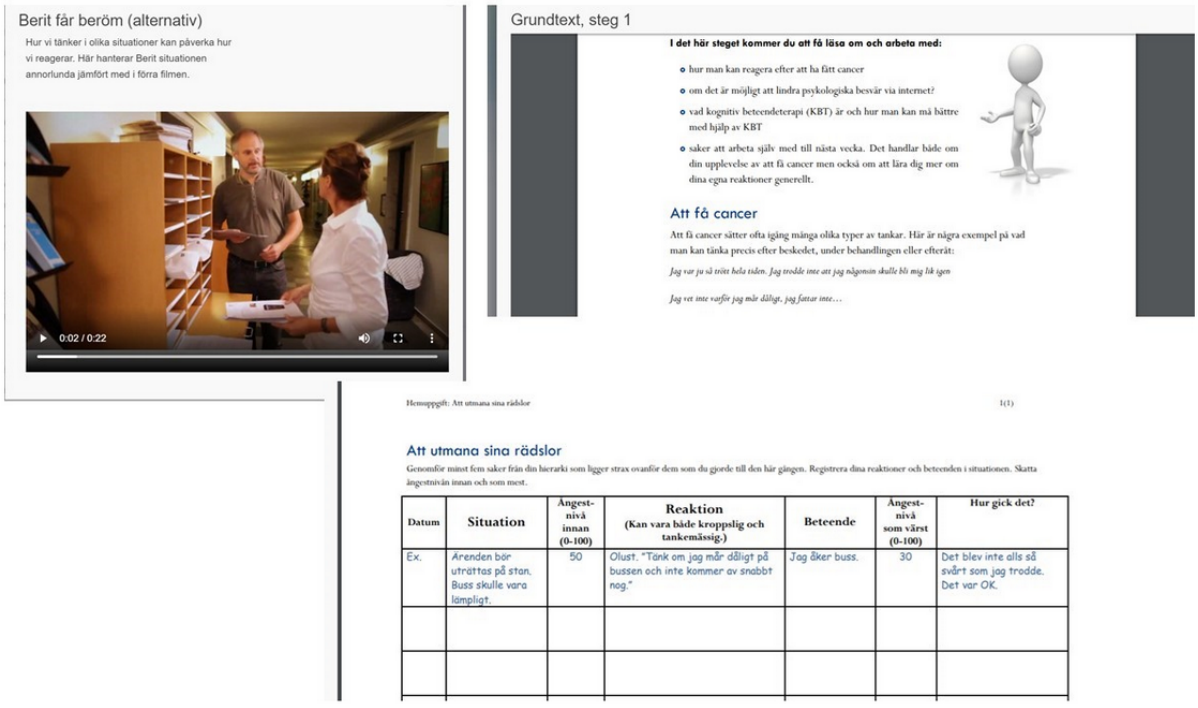


Information was available in text, audiovisual presentations, slideshows, and video clips. All contents were visible to all users, though some parts of the information concerned a specific diagnosis. Step 1 also contained a forum, a chat function, and a frequently asked questions (FAQ) section, alongside the feature *Ask an expert*, where users could pose questions (see Figure 5). In order to strengthen the therapeutic alliance, the nurses (Step 1) and psychologist (Step 2) were presented with a photo and a brief description [16].

Figure 5. Frequently asked questions (FAQ), *Ask an expert*, and chat function (in Swedish).

Step 2 in iCAN-DO constituted a 10-week internet-based cognitive behavioral therapy (iCBT) program [17] managed by a psychologist. iCBT was offered once and only to participants with prevailing symptoms of anxiety or depression (ie, score >7 on either HADS subscale) at 1, 4, or 7 months after inclusion. The treatment comprised 15 modules with written text, audiovisual presentations, and video clips covering a range of problem areas that were relevant for this population, including worry, fatigue, and depressed mood [18]. See Figure 6 for examples of a video clip, text, and homework assignment in iCBT.

Figure 6. Examples of a video clip, text, and homework assignment in internet-based cognitive behavioral therapy (iCBT; in Swedish).



All participants accepting iCBT were offered an introductory module, after which they were free to choose the most relevant modules to work with over 10 weeks. A psychologist guided the participants by monitoring their work, answering any questions, and delivering weekly feedback. All communication took place in the form of written messages via the portal.

Informants

To be included, informants had to have had access to iCAN-DO for at least 7 months. In order to obtain variation among informants, a purposive selection was performed, mainly regarding activity in Step 1 and whether they participated in iCBT or not, but also regarding age, gender, and general online activity. A total of 20 individuals were approached, of whom 2 declined to participate and 3 were unreachable. In total, 15 informants were interviewed (see Table 1) in 2016 and 2017. All informants used the internet on a daily basis, but their general online activity varied. A total of 2 informants (13%) described themselves as very inexperienced computer users.

Procedure

A letter was sent to potential informants with study information. They were informed that they would receive a phone call from research personnel within a couple of days asking about willingness to participate. If they did not want to receive this phone call, they could contact the person responsible for *AdultCan*; no one used that option. When a person consented to participate, a time and place for an interview were booked. All informants were treated with confidentiality and had time to consider their participation after getting written and verbal information. All provided signed informed consent. The study was approved by the regional committee for research ethics in Sweden, Uppsala county (Dnr 2012/003/9).

Data Collection

Questionnaires and Log Data

Background information on informants was retrieved from self-reported questionnaires. All informants' activity in iCAN-DO was logged in the portal. Information about general online activity was retrieved at the time of the interviews.

Interviews

The second author (AH) was the interviewer. AH is a registered specialist nurse with experience in palliative cancer care. AH has been involved in some parts of the development and delivery of the stepped-care intervention, but has no overall responsibility in the program. An interview guide with open-ended questions was used, containing questions regarding participants' user experiences of iCAN-DO. Follow-up questions were based on each informant's responses and were used to acquire more thorough descriptions of experiences [19]. Questions such as "How did that feel?" or "How did you use that?" were used, as well as probing and interpreting questions. One test interview was performed to explore the design and understanding of questions, but no adjustments to the interview guide were needed. Informants were offered the option to have iCAN-DO open during the interview, if they wanted to. The interviews were performed in a place chosen by the informant; some informants chose their place of work, some wanted to meet in their home, and a few chose the hospital. The interviews were tape-recorded and lasted between 45 and 120 minutes.

Data Analysis

Interviews were transcribed verbatim by AH, and latent content analysis [20] was used to analyze the interviews. The text was read several times and the manifest content (ie, what is explicit and obvious in the text) of each interview was divided into meaning units and condensed meaning units. Each condensed meaning unit was then given a code to describe the key message, and codes with similar content were allocated to the same category (see Table 2). An interpretation of the latent content (ie, the underlying meaning of the text) was made by abstracting the underlying implicit aspects of the manifest and explicit content. Thus, when several categories were identified as containing similar, repeating ideas, a subtheme was formed. The implicit underlying meaning of all subthemes was then abstracted to one overarching theme. AH and HI were mainly responsible for the analysis, but the group of authors—mainly HI, AH, BJ, and SA—discussed and considered the results on several occasions. The quotes presented in the paper were translated by the first author (HI), AH, and Linnea Holmén at Calyptic.

Table 1. Informants' characteristics at time of inclusion in the randomized controlled trial (RCT) AdultCan.

Characteristic	Value (N=15)
Age (years), mean (SD); min-max	58.9 (8.9); 37-69
Gender and diagnosis[a], n (%)	
Female	
Total	10 (67)
Breast cancer	10 (67)
Male	
Total	5 (33)
Colorectal cancer	1 (7)
Prostate cancer	4 (27)
Relationship status, n (%)	
Married or partner and living with someone	12 (80)
Married or partner but living alone	1 (7)
Widowed	1 (7)
Single	1 (7)
Level of education, n (%)	
Elementary school	3 (20)
High school	1 (7)
University up to 3 years	6 (40)
University more than 3 years	5 (33)
Working situation, n (%)	
Working	10 (67)
Retired	4 (27)
Early retirement	1 (7)
General online activity (outside the portal), n (%)	
Daily with no social media	4 (27)
Daily and active on social media: lurking	7 (47)
Daily and active on social media: participating	4 (27)
Activity in Step 1, n (%)	
Opening material (all sections[b]) >20 times	6 (40)
Opening material (some sections) >20 times	7 (47)
Opening material (some sections) <10 times	2 (13)
Participation in Step 2 (iCBT[c]), n (%)	
No	9 (60)
Yes	6 (40)

[a]Diagnosis percentages are out of the total number of informants (N=15).

[b]Library, peer support, frequently asked questions, or Ask an expert.

[c]iCBT: internet-based cognitive behavioral therapy.

Table 2. Steps in the content analysis.

Meaning unit	Condensed meaning unit	Code	Category
That you can sit at home, in peace and quiet, when you feel like it and something comes up. You seldom come to think of things when you have the possibility to ask questions, at the hospital.	Studying the material when the questions arise	Available when the need arises	A web portal provides high availability and accessibility
The disease means that it's hard for me to get going, I feel best in the evening and that's when I can gather information in the portal.	Being able to gather information when I feel well	Availability tailored for my health	A web portal provides high availability and accessibility

Results

Overview

In the analysis, 10 categories were derived concerning user experiences and perceptions of delivery, design, and structure of iCAN-DO. The categories were abstracted into three subthemes with the overarching main theme *The cancer disease and its treatment place high demands on accessibility and user experience* (see Table 3).

Subtheme 1: User Experience in the Context of Cancer

The personal situation and consequences of illness, both somatic and cognitive, were mentioned by all informants as factors that in various ways affected their ability to make use of iCAN-DO. Symptoms of fatigue and pain could require being in a comfortable position when listening or reading. Some informants considered it facilitating to be able to use iCAN-DO on tablets as well as on a computer, even though the design was not optimized for mobile devices. Furthermore, having direct access to iCAN-DO was perceived as beneficial by those experiencing a lot of time constraints in everyday life. They could use iCAN-DO when it suited them and when they had the time.

The information in Step 1 was provided in different formats, and informants had varying preferences in regard to the material. Some assimilated the contents best by reading, others by listening to audio clips or watching video clips, and others by using some combination of these formats. Some printed all the material and read it on paper. Though video facilitated immersion in the material, in some cases video clips were considered too long or had poor audio quality, which hampered the informants' ability to pay attention. Furthermore, informants stated that answering questionnaires evoked both negative and positive feelings, but also made them reflect on their health. They felt questionnaires could be a part of health care, for example, as preparation ahead of a visit to oncology care.

Subtheme 2: Technical Struggles Require Patience and Troubleshooting

Many of the informants described the multistep log-in as an overly difficult way to log in, and an option of using Mobile BankID, an electronic identification application, would have been preferred by most. Some of the informants described how having different log-in codes caused problems. The multistep log-in also reduced spontaneous visits. The informants also described other technical struggles, such as the need for separate add-ons to make the page work and a perceived instability, for example, being involuntarily logged out and difficulties when communicating with the psychologist. Some of these problems were described as transient and occurring only in the initial phase, while others were seen as persistent. Perceived technical difficulties were described as a factor that caused negative feelings, lack of motivation, and a postponing behavior.

The prevalence of technical problems required a lot of problem solving. The informants described different ways of handling problems. Some informants described solving problems themselves by installing new software or requesting a new password. Others enlisted the help of relatives or friends. The importance of technical support was highlighted and those who contacted information technology (IT) support were satisfied with the help they got. Some preferred calling and others sent an email or direct message via the portal. Technical support was described as crucial for not giving up.

The informants also had suggestions for improvements to the interface. An aspect described as important was to receive a confirmation that a written message had reached the help desk when awaiting an answer.

Subtheme 3: Appealing and Usable, but Rather Simple

iCAN-DO was experienced as easier to use than many other systems, and getting access to the contents was seen as more important than getting an advanced system. Some informants had the perception of an intuitive interface that was easy to understand and navigate, while others perceived the same simplicity as a shortcoming. The graphical design was described by some as simple and serving its purpose, while others found it dull or overly intricate. The experiences varied between informants who were experienced everyday internet users with high interest in social media and those who were more inexperienced with little or no use of social media.

When participants had questionnaires to answer in the portal, repeated reminders were sent via SMS. Informants described the reminders as a trigger for accessing the supporting material as well, as it meant they logged in. The informants also came up with ideas regarding functions that could improve iCAN-DO, such as better layout of questionnaires, adding news coverage, and adding graphical displays of self-reports.

Table 3. The results of qualitative content analysis—subthemes, categories, and quotes—for the theme The cancer disease and its treatment place high demands on accessibility and user experience.

Subtheme and categories	Quotes
User experience in the context of cancer	
My health status and environment affect how I use the portal	"The tablet is light and the computer is heavy...so if my body is hurting in every fiber, I don't want to have something heavy in my lap, so I used the tablet almost all the time..." [Informant #3]
Multiple delivery modes enable assimilation of the content	"The mini lectures have been good, maybe they could have been a bit shorter...because it's been tough when you're sick to listen for such a long time." [Informant #13]
A web portal provides high availability and accessibility	"Like, when you are on a bus or train you might want to use the time to check something, you have to be able to do that. Right now, I use my smartphone more than my computer; I only use the computer when I really have to." [Informant #7]
Questionnaires make you reflect on your health	"I could feel a little...well, blue, afterwards. You think...Should I feel like that too?" [Informant #2]
	"...even just answering the questions has been therapeutic. I noticed a difference myself too. At the start I indicated how bad I felt and I noticed that it did actually get better." [Informant #12]
Technical struggles require patience and troubleshooting	
Complicated multistep log-in procedure	"It was awfully tricky, and I should really be able to do it, because I've worked a bit with VPN [virtual private network] connections, but this one crashed several times for me; it didn't work..." [Informant #3]
Annoying technical problems reduce motivation to use the portal	"You get a response within the response itself and then I have an inbox as well, where it ended up. So you have it in two places and even though I have read the response, it still looks unread in the inbox." [Informant #9]
	"Technical hassles are not good, it's very annoying...and once you postpone your visit it might never happen." [Informant #10]
Troubleshooting and the importance of support	"Because otherwise, if you don't get help, you feel like you don't give a damn...When there are technical difficulties on the site you postpone your visits, I will log in later...but then you never do it. But then at the end I got information [from the help desk] so..." [Informant #1]
Appealing and usable, but rather simple	
Intuitive but unstable interface	"I thought it was easy to see what it was about, what they had...if there was something that appealed to you, then it was easy to access and I could see what it said. I'm not a computer person, really, but I managed pretty well, I think, and that's a good sign." [Informant #4]
Appealing, yet simple graphical design	"For people in general, it's probably pretty good that it is so simple. Because then it will be used." [Informant #14]
	"I am into Instagram, Facebook, and all those places, you know, so I felt this was...a bit boring. I am sorry to say that, but that's just me. It doesn't give me the kind of input and variety I need." [Informant #11]
Technology that creates motivation and innovation	"It was good to have the reminders, because sometimes you maybe don't have the energy the first time, you're having a bad day, but then maybe you can manage it after the second or third reminder...I have often looked around a bit at other things when I was logged in on the portal anyway." [Informant #11]
	"In that part, you would want a graphical...like a curve so that you can see it yourself...Another thing that I've thought about is that there could be like a news function, things that are going on within oncology care, like in all of Sweden, maybe." [Informant #12]

Discussion

Principal Findings

According to the informants, the user experience of iCAN-DO was affected by physical and cognitive symptoms, as well as time constraints in everyday life, which all highlighted the importance of 24-hour access, a responsive design, and a well-functioning system providing access to technical support.

Both health-related problems and external circumstances may constitute barriers for making contact with health care in the daytime or during working hours. By using iCAN-DO, informants could study information at a time suitable to their situation and health, and could review the information as many times as needed. The delivery of content in different formats also stood out as a positive feature, as informants reported such varied preferences in this regard. In the developmental phase, the use of different information formats was suggested as highly important by patient representatives, who had their own experiences of fatigue or other debilitating symptoms. Further, the participants underlined the benefit of the contents being accessible via different devices, such as computer, tablet, or mobile phone. In the trial, a responsive design was later

incorporated into the software in order to adapt the portal to mobile devices, including smartphones and tablets. Regardless of which functions are incorporated and what devices are used, it is important to include functions for recording portal activity and usage, to secure internal validity [21].

Informants also described various aspects of filling out questionnaires. Some reported that the questionnaires reminded them of their situation in a negative way, but others perceived the questionnaires as enlightening and providing them with feedback on their own health over time. Other studies have reported promising results on individual tailoring of content depending on diagnosis, treatment, and even tracking of symptoms [22], given that the individual does not feel distress at such monitoring [23]. To motivate patients and professionals to use patient-reported symptoms, it is important to underline the aim of this use and to clearly state which professionals have the responsibility to review and discuss the results [24].

Various aspects of technical struggles were revealed in the interviews. Before launching iCAN-DO in *AdultCan*, user-experience testing was performed in various settings, by both professionals and patient representatives. Still, informants in this study described difficulties, and technical support was crucial for not giving up. Experiencing difficulties when attempting to use a program has been described to influence perceptions of its value as a whole. However, training and technical support can make users feel more capable over time [25,26]. This is in line with the Technology Acceptance Model [27], which posits the predictive power of the person's perceived usefulness of the system and ease of use. Though the informants in this study mentioned perceived benefits from the 24/7 access and the multifaceted information, the technical problems when logging in or filling out questionnaires were described as hampering use. Some problems were transitional and could be solved. The possibility to contact a support desk was crucial, suggesting that a highly accessible support function may be very important for new web portals. A take-home message is that electronic health (eHealth) initiatives, to a higher extent, should look into how to provide more user support using technology. For instance, in multiple fields of practice, there is an increasing use of chatbots (ie, text-based conversational interface) to provide customers with support. Such technologies have also been considered in the health care setting when it comes to medical counseling (eg, in oncology care [28]), but could be extended to include technical support to (1) provide a better user experience and (2) increase the scalability of eHealth solutions by reducing the need for support desk staff.

Since iCAN-DO was part of an RCT, an authorization procedure was necessary for control and internal validity. However, many informants perceived the multistep log-in procedure as burdensome, which has also been recognized in similar platforms [29]. After this set of interviews, access to iCAN-DO has been facilitated by the incorporation of Mobile BankID, which drastically reduced the need for support regarding log-in matters. BankID is the leading electronic identification solution in Sweden and has around 8 million active users out of the total population of 10 million people. Even though this particular solution is not available in all countries, we want to stress the need to investigate and adopt electronic identification solutions that are prevalent in the national care provision context and that are well known to the users.

The experiences of the design and the user experience of iCAN-DO varied between the informants who were experienced internet users with a high interest in social media and those with little or no use of social media. As stated earlier, patient representatives from the three diagnostic groups participated in the development of iCAN-DO to increase the relevance of the internet-based stepped care. This group of patient representatives was a quite homogenous group with regard to their opinions about the graphical design, stressing the need for simplicity. It is worth highlighting that more computer experience has been reported to be associated with increased use in another web-based intervention for patients with cancer [30]. In a review of patient web portals for primary and secondary prevention, sociodemographic factors, such as belonging to a racial and ethnic minority, having less education, and having lower health literacy, have been found to be associated with less system use [31]. Thus, when designing future web-based interventions, efforts should be made to do a user analysis and to include patient representatives with varied experience of social media, internet use, education, health literacy, and ethnicity. Furthermore, future web portals not restricted to research projects may benefit from updating the content at regular intervals [28] and incorporating features with news and statistics.

Methodological Considerations

The aim of this study was to evaluate the user experiences of iCAN-DO used in *AdultCan*; thus, the only people available for inclusion were participants in the ongoing trial, which might lower the transferability to other people and contexts. The sample also had a predominance of female informants with breast cancer, reflecting the total sample in the RCT. The majority of the informants were well educated, which may act as a confounder for both perceived usefulness and user experiences of the portal and its material. Beyond differences in diagnoses and education, informants in this study varied in other aspects, such as general online activity and age. Even though the results relate only to this particular portal and web-based support, the take-home message regarding accessibility, delivery mode, technical matters, and design aspects is one that researchers and designers of other web-based support might benefit from. Future studies should consider the user experience in light of not only age, gender, and diagnosis but also factors such as ethnic background, education, and health literacy.

Conclusions

The high accessibility and variety of delivery modes provided by iCAN-DO were important, but consequences of illness, as well as individual situations and preferences, placed high demands on user experience. Technical struggles decrease motivation and usage, which highlights the importance of a support function. Users' internet activity, computer experience, and interactions in social media are aspects that seem to impact on the perception of the interface and graphical design. Future web portals would gain from tailoring both content and design to individual preferences.

Acknowledgments

The study was funded by grants from Forte and the Swedish Cancer Society.

Authors' Contributions

HI participated in data analysis and writing of the manuscript. AH participated in research design, data collection, data analysis, and the writing of the paper. SA was involved in the researcher triangulation process and reading and revision of the manuscript. JS and ÅC contributed to the reading and revision of the manuscript. BJ participated in research design, data analysis, and the writing of the paper.

Conflicts of Interest

None declared.

References

1. DeLisle J. Designing an acuity tool for an ambulatory oncology setting. Clin J Oncol Nurs 2009 Feb;13(1):45-50. [doi: 10.1188/09.CJON.45-50] [Medline: 19193548]
2. Osborn RL, Demoncada AC, Feuerstein M. Psychosocial interventions for depression, anxiety, and quality of life in cancer survivors: Meta-analyses. Int J Psychiatry Med 2006;36(1):13-34. [doi: 10.2190/EUFN-RV1K-Y3TR-FK0L] [Medline: 16927576]
3. Kim SH, Park S, Kim SJ, Hur MH, Lee BG, Han MS. Self-management needs of breast cancer survivors after treatment: Results from a focus group interview. Cancer Nurs 2020;43(1):78-85. [doi: 10.1097/NCC.0000000000000641] [Medline: 30148729]
4. Matsuyama RK, Kuhn LA, Molisani A, Wilson-Genderson MC. Cancer patients' information needs the first nine months after diagnosis. Patient Educ Couns 2013 Jan;90(1):96-102. [doi: 10.1016/j.pec.2012.09.009] [Medline: 23058682]
5. Walsh J, Harrison JD, Young JM, Butow PN, Solomon MJ, Masya L. What are the current barriers to effective cancer care coordination? A qualitative study. BMC Health Serv Res 2010 May 20;10:132 [FREE Full text] [doi: 10.1186/1472-6963-10-132] [Medline: 20482884]
6. Sawyer SM, McNeil R, McCarthy M, Orme L, Thompson K, Drew S, et al. Unmet need for healthcare services in adolescents and young adults with cancer and their parent carers. Support Care Cancer 2017 Jul;25(7):2229-2239. [doi: 10.1007/s00520-017-3630-y] [Medline: 28261754]
7. Murray E, Burns J, See TS, Lai R, Nazareth I. Interactive health communication applications for people with chronic disease. Cochrane Database Syst Rev 2005 Oct 19(4):CD004274. [doi: 10.1002/14651858.CD004274.pub4] [Medline: 16235356]
8. Barr N, Vania D, Randall G, Mulvale G. Impact of information and communication technology on interprofessional collaboration for chronic disease management: A systematic review. J Health Serv Res Policy 2017 Oct;22(4):250-257. [doi: 10.1177/1355819617714292] [Medline: 28587494]
9. Mattsson S, Alfonsson S, Carlsson M, Nygren P, Olsson E, Johansson B. U-CARE: Internet-based stepped care with interactive support and cognitive behavioral therapy for reduction of anxiety and depressive symptoms in cancer--A clinical trial protocol. BMC Cancer 2013 Sep 11;13:414 [FREE Full text] [doi: 10.1186/1471-2407-13-414] [Medline: 24024826]
10. Portz JD, Bayliss EA, Bull S, Boxer RS, Bekelman DB, Gleason K, et al. Using the technology acceptance model to explore user experience, intent to use, and use behavior of a patient portal among older adults with multiple chronic conditions: Descriptive qualitative study. J Med Internet Res 2019 Apr 08;21(4):e11604 [FREE Full text] [doi: 10.2196/11604] [Medline: 30958272]
11. Sjöström J, von Essen L, Grönqvist H. The origin and impact of ideals in eHealth research: Experiences from the U-CARE research environment. JMIR Res Protoc 2014 May 23;3(2):e28 [FREE Full text] [doi: 10.2196/resprot.3202] [Medline: 24860071]
12. Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
13. Hauffman A, Alfonsson S, Mattsson S, Forslund M, Bill-Axelson A, Nygren P, et al. The development of a nurse-led internet-based learning and self-care program for cancer patients with symptoms of anxiety and depression: A part of U-CARE. Cancer Nurs 2017;40(5):E9-E16. [doi: 10.1097/NCC.0000000000000402] [Medline: 27223884]
14. Orem DE. Nursing: Concepts of Practice. 6th edition. St. Louis, MO: Mosby; 2001.
15. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
16. Bisseling E, Cillessen L, Spinhoven P, Schellekens M, Compen F, van der Lee M, et al. Development of the therapeutic alliance and its association with internet-based mindfulness-based cognitive therapy for distressed cancer patients: Secondary

analysis of a multicenter randomized controlled trial. J Med Internet Res 2019 Oct 18;21(10):e14065 [FREE Full text] [doi: 10.2196/14065] [Medline: 31628791]
17. Andersson G. Using the internet to provide cognitive behaviour therapy. Behav Res Ther 2009 Mar;47(3):175-180. [doi: 10.1016/j.brat.2009.01.010] [Medline: 19230862]
18. Harrington CB, Hansen JA, Moskowitz M, Todd BL, Feuerstein M. It's not over when it's over: Long-term symptoms in cancer survivors--A systematic review. Int J Psychiatry Med 2010;40(2):163-181. [doi: 10.2190/PM.40.2.c] [Medline: 20848873]
19. Bolderston A. Conducting a research interview. J Med Imaging Radiat Sci 2012 Mar;43(1):66-76. [doi: 10.1016/j.jmir.2011.12.002] [Medline: 31052024]
20. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: Concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
21. Danaher BG, Seeley JR. Methodological issues in research on web-based behavioral interventions. Ann Behav Med 2009 Aug;38(1):28-39 [FREE Full text] [doi: 10.1007/s12160-009-9129-0] [Medline: 19806416]
22. Langius-Eklöf A, Christiansen M, Lindström V, Blomberg K, Hälleberg Nyman M, Wengström Y, et al. Adherence to report and patient perception of an interactive app for managing symptoms during radiotherapy for prostate cancer: Descriptive study of logged and interview data. JMIR Cancer 2017 Oct 31;3(2):e18 [FREE Full text] [doi: 10.2196/cancer.7599] [Medline: 29089290]
23. Lupton D. The digitally engaged patient: Self-monitoring and self-care in the digital health era. Soc Theory Health 2013 Jun 19;11(3):256-270. [doi: 10.1057/sth.2013.10]
24. Kuijpers W, Groen WG, Loos R, Oldenburg HS, Wouters MW, Aaronson NK, et al. An interactive portal to empower cancer survivors: A qualitative study on user expectations. Support Care Cancer 2015 Sep;23(9):2535-2542 [FREE Full text] [doi: 10.1007/s00520-015-2605-0] [Medline: 25876157]
25. Grindrod KA, Li M, Gates A. Evaluating user perceptions of mobile medication management applications with older adults: A usability study. JMIR Mhealth Uhealth 2014 Mar 14;2(1):e11 [FREE Full text] [doi: 10.2196/mhealth.3048] [Medline: 25099993]
26. Czaja S. The potential role of technology in supporting older adults. Public Policy Aging Rep 2017;27(2):44-48. [doi: 10.1093/ppar/prx006]
27. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
28. Bibault J, Chaix B, Guillemassé A, Cousin S, Escande A, Perrin M, et al. A chatbot versus physicians to provide information for patients with breast cancer: Blind, randomized controlled noninferiority trial. J Med Internet Res 2019 Nov 27;21(11):e15787 [FREE Full text] [doi: 10.2196/15787] [Medline: 31774408]
29. Groen WG, Kuijpers W, Oldenburg HS, Wouters MW, Aaronson NK, van Harten WH. Supporting lung cancer patients with an interactive patient portal: Feasibility study. JMIR Cancer 2017 Aug 08;3(2):e10 [FREE Full text] [doi: 10.2196/cancer.7443] [Medline: 28790025]
30. Børøsund E, Cvancarova M, Ekstedt M, Moore SM, Ruland CM. How user characteristics affect use patterns in web-based illness management support for patients with breast and prostate cancer. J Med Internet Res 2013 Mar 01;15(3):e34 [FREE Full text] [doi: 10.2196/jmir.2285] [Medline: 23454601]
31. Coughlin SS, Prochaska JJ, Williams LB, Besenyi GM, Heboyan V, Goggans DS, et al. Patient web portals, disease management, and primary prevention. Risk Manag Healthc Policy 2017;10:33-40 [FREE Full text] [doi: 10.2147/RMHP.S130431] [Medline: 28435342]

Abbreviations

eHealth: electronic health
FAQ: frequently asked questions
HADS: Hospital Anxiety and Depression Scale
iCBT: internet-based cognitive behavioral therapy
IT: information technology
RCT: randomized controlled trial
U-CARE: Uppsala University Psychosocial Care Program

Edited by G Eysenbach; submitted 09.10.19; peer-reviewed by T Freeman, M Christiansen; comments to author 15.01.20; revised version received 04.03.20; accepted 21.03.20; published 19.05.20.

Please cite as:
Igelström H, Hauffman A, Alfonsson S, Sjöström J, Cajander Å, Johansson B
User Experiences of an Internet-Based Stepped-Care Intervention for Individuals With Cancer and Concurrent Symptoms of Anxiety or Depression (the U-CARE AdultCan Trial): Qualitative Study
J Med Internet Res 2020;22(5):e16604
URL: http://www.jmir.org/2020/5/e16604/
doi:10.2196/16604
PMID:32427108

Original Paper

Help to Overcome Problems Effectively for Cancer Survivors: Development and Evaluation of a Digital Self-Management Program

Faith Martin[1], PhD, DClinPsy; Hayley Wright[2*], BSc, MSc, PhD; Louise Moody[3*], BSc, PhD; Becky Whiteman[4,5*], RN, MA; Michael McGillion[6*], BScN, PhD; Wendy Clyne[7,8*], PhD; Gemma Pearce[9*], PhD; Andy Turner[2*], BSc, PhD

[1]Faculty of Health and Applied Science, University of the West of England, Bristol, United Kingdom
[2]Faculty Research Centre for Intelligent Healthcare, Faculty of Health and Life Sciences, Coventry University, Coventry, United Kingdom
[3]Faculty Research Centre for Arts, Memory and Communities, Faculty of Arts and Humanities, Coventry University, Coventry, United Kingdom
[4]UK Early Cancer Detection Consortium, Faculty of Health and Life Sciences, Coventry University, Coventry, United Kingdom
[5]Baxter Healthcare, Compton, Newbury, United Kingdom
[6]School of Nursing, McMaster University, Hamilton, ON, Canada
[7]Hope for The Community, Community Interest Company, The Enterprise Hub, Coventry, United Kingdom
[8]National Institute for Health Research, Research Design Service South West, Peninsula Medical School, Plymouth University, Devon, United Kingdom
[9]Faculty of Health and Life Sciences, School of Psychological, Social and Behavioural Sciences, Coventry University, Coventry, United Kingdom
[*]these authors contributed equally

Corresponding Author:
Andy Turner, BSc, PhD
Faculty Research Centre for Intelligent Healthcare
Faculty of Health and Life Sciences
Coventry University
Priory Street
Coventry, CV1 5FB
United Kingdom
Phone: 44 02477657498
Email: hsx116@coventry.ac.uk

Abstract

Background: People living with cancer face numerous psychosocial challenges, including cancer-related fatigue, fear of recurrence, and depression. There is a lack of digital interventions tailored to the needs of people living with all types of cancer. We developed a 6-week, digital, peer-delivered, self-management program: iHOPE (Help to Overcome Problems Effectively; where 'i' indicates the digital version of the program). The program is underpinned by positive psychology and cognitive behavioral therapy to meet these psychosocial challenges.

Objective: This study aimed to assess the feasibility of the iHOPE program among people living with cancer. Program adherence and satisfaction along with changes in psychological distress and positive well-being were measured.

Methods: A pre-post, acceptability, and feasibility design was used. People living with cancer (N=114) were recruited via a national cancer charity in the United Kingdom and were given access to the iHOPE program. Demographic and other participant characteristics were recorded. Participants completed digital measures at baseline and the end of the 6-week program for depression, anxiety, cancer-related fatigue, cancer *worry* or fear of cancer recurrence, positive mental well-being, hope, gratitude, and health status. The website's system recorded data on the usage of the program. Satisfaction with the program was also measured.

Results: A total of 114 participants completed the baseline questionnaires. Of these, 70 people (61.4%) participated in all 6 sessions. The mean number of sessions undertaken was 5.0 (SD 1.5). Moreover, 44.7% (51/114) of participants completed at least three sessions and end-of-program outcome measures. A total of 59 participants completed the satisfaction questionnaire, where ≥90% (54/58) of participants reported that the program was easy to navigate and was well managed by the peer facilitators, and that they found the social networking tools useful. Preliminary efficacy testing among the 51 participants who completed baseline and postprogram outcome measures showed that postprogram scores decreased for depression, anxiety, cancer-related

fatigue, and fear of recurrence (all $P<.001$) and increased for positive mental well-being ($P<.001$), hope (both $P<.001$), and gratitude ($P=.02$).

Conclusions: The feasibility evidence is promising, showing that the peer-delivered digital iHOPE program is acceptable and practical. Implementation of the iHOPE program on a wider scale will incorporate further research and development to maximize the completion rates of the measures. Initial effectiveness data suggest positive impacts on important cancer-related quality of life and mental well-being outcomes. A randomized controlled trial design with a longer follow-up is needed to confirm the potential of the iHOPE program for improving mental and physical health outcomes for cancer survivors.

(J Med Internet Res 2020;22(5):e17824) doi:10.2196/17824

KEYWORDS

positive psychology; self-management; hope; quality of life; survivorship; cancer

Introduction

Background

Globally, there were an estimated 18.1 million new cases of cancer in 2018 [1]. The worldwide incidence of cancer is predicted to rise by 75% over the next two decades [2], leading to a world cancer burden of around 25 million cases by 2030 [3]. Globally, there are around 43.8 million people living 5 years beyond their diagnosis [1]. On the basis of the most recently collected national datasets in the United Kingdom, 4.7 million new cases of cancer were detected in 2018, with 13.5 million people living 5 years beyond their diagnosis [4]. The number of cancer survivors in the United Kingdom alone is projected to increase by approximately 1 million per decade from 2010 to 2040 [5]. Combined with effective treatments, this leads to a growing population of cancer survivors, many with unmet needs, and experiencing psychosocial and physical difficulties.

Cancer survivors face a number of challenges following primary treatment, including fatigue, pain, sexual problems, cognitive functioning, fear of cancer recurrence, depression, anxiety, social isolation, and financial issues [6-8]. A substantial number of studies report such difficulties in the long term, particularly when treatment ends and contact with health care professionals diminishes, patients often report feeling abandoned, vulnerable, and as if they have lost the *safety net* they felt they had during treatment [7]. Clinically significant cancer-related fatigue is common [9,10]. Research suggests that after treatment has ended, a significant proportion of cancer patients experience fear of recurrence [11], with potential long-term negative impacts on quality of life and mental health, including hypervigilance, anxiety, posttraumatic stress, and depression [11-16]. Support for fear of cancer recurrence is an unmet need for between 22% and 90% of cancer survivors [12]. Overall, 12% to 20% of survivors of cancer meet diagnostic criteria for major depression and 18% to 40% for an anxiety disorder in the first 2 years following diagnosis [15,16]. Although these problems are frequently encountered in clinical practice, no clear consensus exists on the best management strategies to support people experiencing anxiety and depression posttreatment.

In the United Kingdom, the National Health Service (NHS) national health care strategy has identified the important role that technology can play in supporting patients with long-term conditions to be better able to self-manage their health [17,18]. This builds on the UK National Cancer Survivorship Initiative, which highlighted the need for a greater focus on recovery, health, and well-being after cancer treatment [19]. Digital self-management interventions are then a central feature of future plans to support cancer survivors.

Self-management interventions have been found to improve outcomes, including quality of life and health care utilization in long-term conditions [20]. Although there are substantial issues concerning intervention fidelity and content, implementation, and scaling and reach of intervention delivery, findings from a review and meta-analysis of digital self-management interventions for cancer survivors found generally positive, support for this approach [21]. A recent systematic review of all cancer self-management interventions highlighted the enormous diversity in intervention content, rendering it difficult to draw conclusions about their effectiveness, and found that there was very poor sustainability of interventions [22]. Furthermore, it is vital to note that many of these research-based interventions are not openly available to the public, with poor implementation observed [23]; indeed, there are very few self-management programs actually available in the United Kingdom [24]. Partnering with implementation stakeholders at an early stage in intervention development is advised to address this challenge [25].

In response to the shortage of available, tailored self-management support programs for cancer survivors, we worked with cancer survivors, clinicians, and other experts to develop a group-based, face-to-face self-management program: Help to Overcome Problems Effectively, known as the *HOPE* program, for survivors of all types of cancer. A recent systematic review of self-management interventions showed that content was largely based on expert opinions or previous models of self‐management or chronic care, with patient input into the design reported in only about 10% of the studies [26]. Therefore, the involvement of cancer survivors in the development, testing, and facilitation of the HOPE program is a particular strength of the intervention compared with other cancer self-management programs. Furthermore, the HOPE program is novel and distinct from many other cancer self-management programs because of its roots in positive psychology [27-29] and its unique focus on hope and gratitude to create an upward spiral of positivity [30] to improve well-being and coping. Fredrickson [31] shows that increasing positive emotions broadens attention, thinking, and action, which enables people to develop more creative thought and action pathways (eg, expanding coping skills), and thus

develop crucial personal and social resources for self-management.

Hope theory [32,33] is similar to self-efficacy theory, but the latter focuses on specific goals and behaviors, whereas hope theory recognizes enduring cross-situational goals and behaviors, and as such, hope theory is better suited to the complexity of managing the diverse impact of long-term conditions. Goals are fundamental in hope theory, which encompasses a cognitive set that is based on both agency (goal-directed determination) and pathways (planning ways of achieving goals).

Gratitude has been shown to improve psychological well-being and increase positive emotions [34,35], with some interventions showing that increasing gratitude is linked to improvements in depression [36]. A gratitude activity is a weekly feature in the HOPE program and is designed to increase participants' positive emotions. The HOPE program also includes other evidence-based cognitive behavioral therapy and positive psychological activities such as identifying personal strengths, scheduling pleasant activities, mindfulness, relaxation training, and reviewing successes.

Group curative factors of instilling hope, universality, and altruism [37] are embedded within the HOPE program content, where participants observe their peers overcoming challenges and achieving goals (*instillation of hope*), share experiences (*universality*), and provide informational and emotional support for each other (*altruism*). Thus, drawing on the principles of positive psychology, the hope theory and gratitude, and embedding group curative factors, the HOPE program provides participants with a novel toolkit to develop skills and resources to improve their well-being and quality of life during and beyond treatment.

Objectives

The HOPE program recognizes the common challenges and unmet needs across all types of cancer, including fatigue, fear of recurrence, and psychological distress [6-16]. The HOPE program was co-designed with service users and stakeholders from one of the United Kingdom's leading cancer charities, Macmillan Cancer Support (MCS) [38,39]. It has been delivered face-to-face to breast cancer survivors and in community settings for all cancer types [38,40]. Initial evaluations demonstrate that cancer survivors participating in the face-to-face program feel more confident and hopeful and valued the peer support element of the program, which made them feel less alone with their problems [38].

MCS experienced difficulties recruiting cancer survivors to the face-to-face program in some regions of the United Kingdom. This mirrors the national and international self-management experience [41]. Digital delivery of interventions can improve access and increase user choice for those who may be unable to physically attend face-to-face programs and those who may prefer a remote, digital intervention [42].

It is unclear if digital interventions can offer the same active intervention ingredients, known as group curative factors [37], that participating in a face-to-face group provides. The growth in social networking potentially offers a strong sense of community, a place for sharing experiences of cancer, and a useful platform for self-management and support [43]. Evidence shows that digital interventions have a positive impact on the quality of life and other health, self-management, and behavioral outcomes in cancer survivors and other long-term conditions [44,45]. Digital interventions also appear to address barriers to participation [46] (eg, physical, psychological, cognitive, economic, social, and cultural factors) in existing face-to-face programs [47]. Therefore, we adapted the existing face-to-face format of the HOPE program to create the digital iHOPE (Help to Overcome Problems Effectively; where the 'i' indicates the digital version to distinguish it from the face-to-face HOPE program) program. In this study, we aimed to explore the feasibility of the digital version of the program available for people living with all types of cancer. As a feasibility study, the specific aims were to examine the following aspects:

- Implementation: recruitment rates, completion rates of measures, and adherence rates to the program (ie, number of participants who participated in the sessions)
- Acceptability: satisfaction ratings, ratings of appropriateness and potential usefulness of the program, and indications of positive and negative effects on participants
- Practicality: feedback on the ability to complete activities in terms of time constraints and other personal commitments
- Preliminary efficacy testing: outcomes and effect size estimation [48-50].

Methods

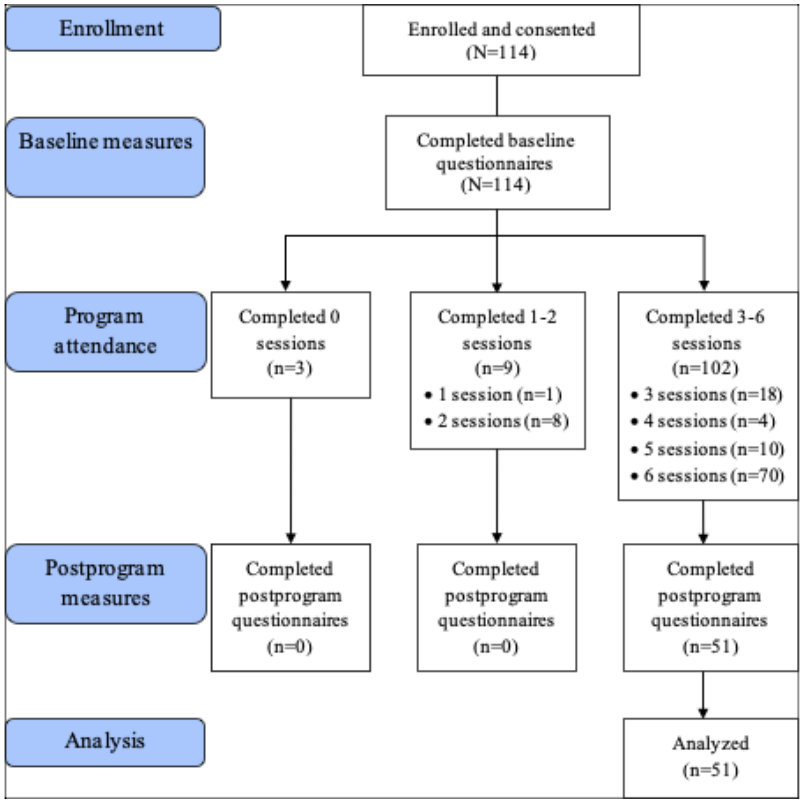
Informed Consent

The study was approved by the Coventry University Ethics Committee (P21296). All participants provided informed consent before participation in the study.

Participants and Recruitment

A convenience sample of cancer survivors recruited to 5 iHOPE programs over 10 months was used (n=114). Eligible study participants were adults (≥18 years), living in the United Kingdom, coming to the end of cancer treatment or surgery, or having recently completed treatment, able to read and understand English, and with access to the internet and an email account. Participants were recruited through advertisement of the program on the MCS Facebook page, Twitter, and Macmillan's *LearnZone* area on their website. Participants were allocated a place on the program on a first-come, first-served basis. In light of previous digital intervention research showing that participants who completed a minimum of 50% of sessions had a reliable change in outcome measures [51], we categorized participants as having completed the course if they completed three or more of the 6 sessions (n=102). All participants completed the baseline questionnaires (n=114), but only those who also completed the postprogram questionnaires were included in the analysis (n=51). Of the remaining 63 participants who did not complete the postprogram questionnaires, 3 participants did not use the program at all, 9 participants completed 1 or 2 sessions, and 51 participants completed between 3 and 6 sessions (see Figure 1).

Figure 1. Participant flow.



Program Development

Full details of the development of the HOPE program have been published elsewhere [38-40,52,53]. The HOPE program has been taxonomized using the taxonomy of self-management support [54]. The adaptation to the digital iHOPE program was undertaken in consultation with people who had attended, delivered, and commissioned the delivery of the face-to-face HOPE program. A user-centered, iterative approach was undertaken [55], as detailed below. A set of design requirements and a design brief were drawn up in consultation with end users and stakeholders. It was specified that the digital version (iHOPE) should replicate the process and content of the group-based HOPE course to ensure that the marketing and recruitment of both versions of the course would be consistent and that cross-training of face-to-face facilitators to deliver iHOPE would be kept to a minimum. The initial digital version of HOPE went through a number of iterative testing sessions, with improvements made to usability after each iteration. It was intended through these iterations to develop a system that was usable and accepted by the intended user group to increase the likelihood of uptake and continued usage, and ensure the technology did not prove a barrier to engagement and participation.

Iteration 1

We conducted usability testing of iHOPE with stakeholders, including MCS staff, trainers, and cancer survivors. A link to the course was circulated to past participants and facilitators of the face-to-face program. They were asked to work through the web-based material and provide feedback to specific questions via email. This study aimed to explore if the core components of HOPE were integrated effectively.

Iteration 2

The iHOPE web-based course was then reviewed by a wider audience and demonstrated to delegates at the National Cancer Voices conference in November 2013. This round of feedback focused on the acceptability of the translation of face-to-face course features into a web format.

Iteration 3

MCS and Coventry University researchers held a workshop with the web design team and experienced HOPE facilitators to collect feedback on iHOPE design, usability, and content. User and facilitator feedback led to further revisions to improve usability and course experience.

Iteration 4

The final iterative feedback was undertaken as part of the iHOPE evaluation, and 5 cohorts reviewed the system while they were enrolled in the program.

Content

The iHOPE program content comprises text, images, downloadable documents, and links to external websites, for example, activities and media related to cancer-related fatigue and developing character strengths. The content delivered is configured into interactive activities (eg, quizzes, self-monitoring tools, and diaries) that can be used by participants to learn and consolidate program content (Table 1 gives details of iHOPE program content and activities, and Figure 2 shows screenshots of user dashboards). The iHOPE program uses forums and messaging facilities that act as a conduit for communication between participants and facilitators.

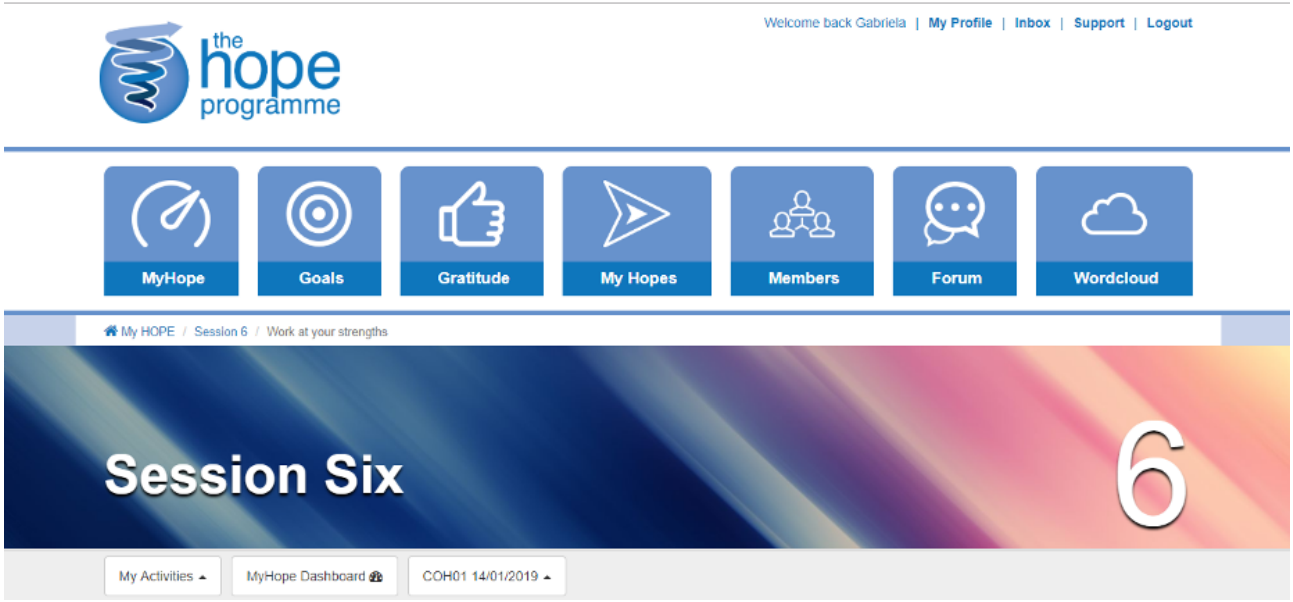
Table 1. Weekly topics, content, exercises, and activities included in the 6-week Help to Overcome Problems Effectively (where 'i' indicates digital version) program.

Session	Examples of content	Examples of exercises and activities (self-management tools)
Week 1: Introduction/instilling hope	• Aims of the program • User guide to navigating the platform and setting up a profile • Introduction to self-management • The benefits of positive emotions • Video: positive emotions for a flourishing life • The power of gratitude • Personalized goal setting • Video: how to set achievable goals • Forum topic: reasons for joining the program • Further resources and links (eg, videos, podcasts, and websites) to gratitude, positivity, and goal setting	• Interactive gratitude diary • SMARTER[a] goal setting • Assessment: positivity ratio test and positive and negative emotions test
Week 2: Stress management	• Understanding stress • Managing stress • Videos: how to manage stress and how to make stress your friend • Coping with unhelpful thinking patterns • Mindfulness for stress management and meditation • Self-compassion and acceptance • Video: how to be kind to yourself • Forum topic: how do you deal with cancer-related stress? • Further resources and links (eg, videos, podcasts, and websites) to self-compassion, mindfulness, and stress management	• Interactive gratitude diary • SMARTER goal setting and goal feedback • Guided relaxation and meditation exercise (podcasts) • How to cope with unhelpful thoughts (worksheet)
Week 3: Managing fatigue	• Understanding the boom and bust cycle • Using the 3 Ps (prioritizing, planning, and pacing) for managing fatigue • Video: tips for managing fatigue • Sleeping better; podcast: tips to improve sleep • Forum topic: coping with fatigue • Further resources and links (eg, videos, podcasts, and websites) to sleeping better	• Interactive gratitude diary • SMARTER goal setting and goal feedback • Fatigue and pacing diaries (worksheets) • Quiz: What are the main challenges faced by cancer survivors?
Week 4: Body image and communication	• Body image • Video: body image and cancer • Sexuality and intimacy • Video: Cancer as a passport to emotional intimacy • Communication skills and tips for talking with the health care team and family • Forum topic: experiences of coping with body changes and experiences of communicating with the health care team • Further resources and links (eg, videos, podcasts, and websites) to sexuality, intimacy, and relationships	• Interactive gratitude diary • SMARTER goal setting and goal feedback
Week 5: Physical activity and fear of recurrence	• Coping with fear of recurrence • Videos: Moving forward while being worried about cancer returning and the regrets of those who are dying • Hopes and dreams for the future • Video: Before I die project • The benefits of physical activity • Video: Tips for becoming and staying active • Forum topic: Concerns about cancer coming back • Further resources and links (eg, videos, podcasts, and websites) to managing concerns about cancer coming back and getting more active	• Interactive gratitude diary • SMARTER goal setting and goal feedback

Session	Examples of content	Examples of exercises and activities (self-management tools)
Week 6: Character strengths and happiness	• Understanding how using your strengths can lead to a more fulfilling life • Video: The science of character strengths • Tips for authentic happiness; managing setbacks and keeping going • Forum topic: Learning from the program • Further resources and links (eg, videos, podcasts, and websites) to Macmillan Cancer Support online communities and happiness resources	• Interactive gratitude diary • SMARTER goal setting and goal feedback • Assessment: positivity ratio test and positive and negative emotions test and character strengths • Quiz: What contributes to happiness?

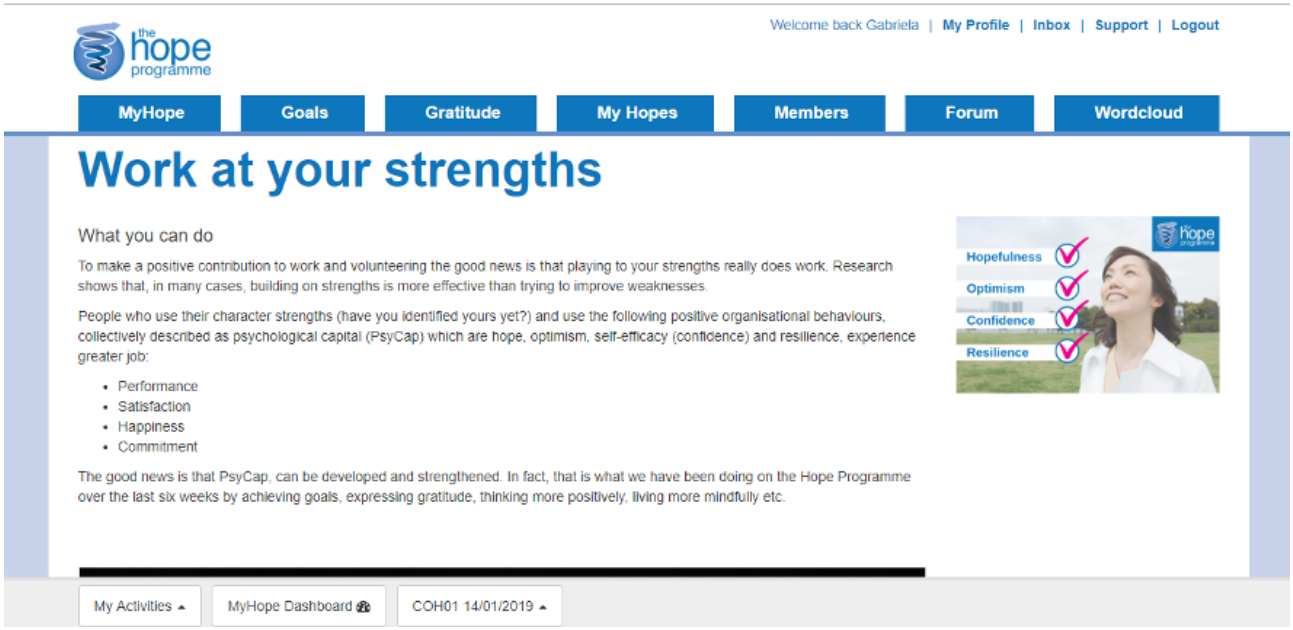
[a]SMARTER: SMARTER is an acronym used by many organizations for goal setting, and stands for: Specific, Measurable, Achievable, Relevant, Time-bound, Enjoyable, Reward.

Figure 2. Help to Overcome Problems Effectively user dashboard.



The theory underpinning the HOPE program is described elsewhere [38,53]. Briefly, the iHOPE program aims to enhance well-being by fostering positive emotions and stimulating positive functioning. A parallel goal is to reduce depressive symptoms. The iHOPE program is based on principles derived from positive psychology and focuses on positive experiences, strengths, and personal competencies rather than mental health problems such as anxiety and depression. It incorporates evidence-based exercises based on positive psychology, in addition to elements stemming from mindfulness, cognitive behavioral therapy, and problem-solving therapy (Figure 3 gives an example of personal strength exercises). The use of positive psychology interventions in cancer survivorship has, to date, been limited; however, a systematic review has supported their impact on well-being and quality of life [56].

Figure 3. Help to Overcome Problems Effectively personal strengths exercise.



Delivery

The maximum group size for each program delivery was set at 20 participants, following consultation with the iHOPE program facilitators and a beta test involving 33 participants. The iHOPE program content is released weekly over 6 weeks, thus mirroring the delivery model of the group-based face-to-face HOPE program. The iHOPE program is an asynchronous program that does not require real-time attendance. On the same day each week, new content is released, whereas previous content remains available. Participants are encouraged to log on for approximately 2.5 hours per week and use a range of behavior change techniques, including weekly goal setting, action planning, and self-monitoring. Peer support and interaction is facilitated through social networking tools and shared interactive activities where all participants' comments appear. Weekly topics and activities are provided in Table 1.

Goal setting and gratitude activities are recurring weekly features, and participants are encouraged to post a goal or something they may feel grateful for on the *online walls* for everyone to see and comment on. Each week, a number of questions are set as discussion topics that are featured in weekly forums. These questions are related to the program content for each week. There is also a weekly forum (Hope Lounge) where participants can start a discussion about either their experience of living with and beyond cancer or a noncancer related topic.

The iHOPE program is moderated by 2 trained peer facilitators who are affected by cancer in some way. The facilitators received training from MCS and followed a delivery protocol. The facilitator's role is to offer encouragement to participants, stimulate discussion in social networking forums by inviting participants to respond with comments to specific questions, or respond to questions/comments posted by participants. Facilitators also monitor daily social networking posts for safety and report any technical problems to the research team. Facilitators spent 2 hours each per session, supporting the participants.

Procedure

One week before each group's iHOPE program start date, participants were sent a link to a web-based survey to access an introductory letter that explained the purpose of the study, the research participant information sheet, consent form, demographic questionnaire, and outcome measures (ie, baseline questionnaires). Participants completed these and then went on to access the iHOPE program. At the end of the 6-week program, participants were emailed the survey link to complete the outcome measures again (ie, postprogram questionnaires) and complete a usability questionnaire.

Data Collection

Data collection was designed to address each of the feasibility testing research aims. Demographic data were collected at baseline on participants' age, sex, ethnicity, marital status, employment status, changes in work hours because of cancer, educational level, and cancer site.

Implementation

Recruitment, retention, and completion rates of the measures were recorded. The iHOPE program system captured basic adherence data, in terms of how many participants accessed the program and the number of sessions they accessed. Furthermore, detailed usage data were captured by the number of times participants carried out key program activities, such as goal setting and gratitude diaries.

Acceptability and Practicality

We created our own bespoke questionnaire to derive satisfaction scores with specific elements of interest for program development purposes, such as navigation, presentation, the usefulness of specific program elements, and support from the program facilitators. As previous literature shows that the optimal number of potential responses on a scale is between 4 and 7 [45], we presented participants with a choice of 4 responses (ie, strongly disagree, disagree, agree, and strongly agree) in an effort to reduce participant response burden. In

addition, as these responses were used to inform program development and changes to future versions, the inclusion of a *neutral* response on the scale would be uninformative. Satisfaction with the program was assessed using a 4-point scale [57] anchored at strongly agree and strongly disagree. The 13 questions (summarized in the Results section and Multimedia Appendix 1) were derived from published research examining the usability of digital self-management interventions [58-60].

Outcome Measures for Preliminary Efficacy Testing

All of the following outcome measures were administered at baseline and again at the end of the iHOPE program.

The quality of life in adult cancer survivors scale (QLACS) [61] is a validated questionnaire comprising 47 items across 12 domains, where 7 domains measure the generic quality of life and 5 domains measure the cancer-specific QLACS. In this study, we used 2 subscales of the QLACS that specifically address frequently cited unmet needs for survivors: cancer-related fatigue and cancer-related concern or fear of recurrence. Fatigue is assessed using the following 4 items: *In the past four weeks... (1) you had the energy to do the things you wanted to do [note that this item was reverse-scored]; (2) you felt fatigued; (3) you did not have energy to do the things you wanted to do; and (4) you felt tired a lot*. Fear of cancer recurrence is measured by the following 4 items: *In the past four weeks… (1) you worried about dying from cancer; (2) you worried about cancer coming back; (3) whenever you felt a pain, you worried that it might be cancer again; and (4) you were preoccupied with concerns about cancer*. Each item in these 4-item subscales is scored from 1 to 7 (1=never, 2=seldom, 3=sometimes, 4=about as often as not, 5=frequently, 6=very often, and 7=always). Items are summed to give an overall score of 4 to 28 for each of the 2 subscales (ie, fatigue and fear of recurrence), with higher scores indicating greater difficulties.

The patient health questionnaire-9 [62] is a 9-item measure that assesses the frequency of experience of the symptoms of depression, for example, *Over the past two weeks, how often have you been bothered by any of the following problems... (1) little interest or pleasure in doing things; (2) feeling down, depressed, or hopeless; and (3) poor appetite or overeating*. Responses to each of the 9 items range from 0 to 3 (0=not at all, 1=several days, 2=more than half the days, and 3=nearly every day), leading to a summed score between 0 and 27, with higher scores indicating greater severity of depression. Scores of 10 or more are presumed to be above the clinical range, and so, scores of less than 10 are classified as *cases* of depression. Recovery rates were calculated as those patients who scored ≥10 (cases) before treatment and scored <10 posttreatment.

The generalized anxiety disorder scale [63] is a 7-item scale measuring symptoms of generalized anxiety disorder, for example, *Over the past two weeks, how often have you been bothered by the following problems… (1) feeling nervous, anxious, or on edge; (2) trouble relaxing; and (3) becoming easily annoyed or irritable*. Responses to all 7 items range from 0 to 3 (0=not at all, 1=several days, 2=more than half the days, and 3=nearly every day), providing a total score of 0 to 21, with higher scores indicating greater anxiety. Scores ≥8 are classified as *cases* of generalized anxiety disorder. Recovery rates were calculated for those patients who score ≥8 (cases) before treatment and <8 post-treatment.

The Warwick Edinburgh mental well-being scale [64] is a scale of 14 positively worded feelings and thoughts, used to assess mental well-being within the adult population. The scale includes measures of positive affect, satisfying interpersonal relationships, and positive functioning, for example, *Below are some statements about feelings and thoughts. Please tick the box that best describes your experience of each over the last two weeks… (1) I have been feeling optimistic about the future; (2) I have been thinking clearly; and (3) I have been feeling love*. Participants rated each of the 14 items on a scale of 1 to 5 (1=none of the time, 2=rarely, 3=some of the time, 4=often, and 5=all of the time), providing a total positive mental well-being score ranging from 14 to 70, with higher scores representing greater positive mental well-being. A change of three or more is seen as clinically *meaningful* change [65].

Hope was measured using the 6-item adult state hope scale [66], which assesses goal-directed thinking in any given situation. The scale has 3 agency items, for example, *At the present time, I am energetically pursuing my goals*, and 3 pathway items, for example, *I can think of many ways to reach my current goals*. Participants indicated the extent to which they agree with each of the 6 statements, in accordance with how they feel at the present moment, on a scale of 1 to 8 (1=definitely false, 2=mostly false, 3=somewhat false, 4=slightly false, 5=slightly true, 6=somewhat true, 7=mostly true, and 8=definitely true). Total hope scores range from 6 to 48, with higher scores indicating higher levels of hopeful thinking.

The gratitude questionnaire-6-item form [67] is a self-report measure of disposition to experience gratitude in everyday life. It is a 6-item scale, comprising items such as *(1) I have so much in life to be thankful for, (2) I am grateful to a wide variety of people, and (3) when I look at the world, I do not see much to be grateful for [note that this item is reverse-scored]*. Participants scored each item from 1 to 7 (1=strongly disagree, 2=disagree, 3=slightly disagree, 4=neutral, 5=slightly agree, 6=agree, 7=strongly agree), giving a total score of 6 to 42, where a higher score indicates more gratitude shown [67].

Analysis

All statistical data analyses were conducted using IBM SPSS Statistics 25 (IBM Corp Released 2017. IBM SPSS Statistics for Windows, Version 25.0). The sample was not powered to detect significance in the outcome measures; nevertheless, we present changes in the scores to aid understanding of the potential effect of the program and to provide data on which to base a power calculation for a larger study of efficacy. The level of statistical significance was set at $P<.05$. Owing to some deviations from the normal distribution in follow-up gratitude, anxiety, and depression scores, all variables were analyzed using Wilcoxon signed-ranks for paired samples. We report the means with SDs and medians with a range of scores. As recommended for nonparametric paired data, we report r as the effect size estimates ($r=z/\sqrt{n}$), where z is calculated through the Wilcoxon test and n is the number of observations (n=102) [68]. Recommended boundaries for r were used to determine small (0.1), moderate (0.3), and large effect sizes (0.5) [69].

Results

Descriptive Data

A total of 114 participants completed the baseline questionnaires, and those who also completed the postprogram questionnaires were included in the analysis (n=51). Participant characteristics are shown in Table 2. Participants were predominately female, white ethnicity, married, with educational qualifications, and employed. The majority had cut work hours because of their cancer, and 38.6% (44/114) had breast cancer; however, data on cancer site was incomplete as participants did not always share this. There were no significant differences between these subgroups of participants on any of the demographic variables.

Examining baseline depression, anxiety, well-being, quality of life, cancer-related fatigue and worry, hope, and gratitude scores between those who did and did not complete postprogram questionnaires revealed no significant differences (scores are summarized in Multimedia Appendix 2).

Table 2. Participant demographic information at enrolment, baseline, and postprogram.

Characteristics	Sample at enrolment (N=114)	Completed baseline questionnaire only (n=63)	Completed baseline and postprogram questionnaire (n=51)
Age, mean (SD)	51.3 (9.6)[a]	47.4 (8.8)[b]	53.6 (9.4)[c]
Female, n (%)	102 (89.5)	56 (89)	46 (90)
White ethnicity, n (%)	111 (97.4)	61 (97)	50 (98)
Married or living with partner, n (%)	86 (75.4)	46 (73)	40 (78)
Employed, n (%)	68 (59.6)	36 (57)	32 (63)
Cut work hours because of cancer, n (%)	73 (64.0)	43 (68)	30 (59)
Possessed postschool qualifications, n (%)	110 (96.5)	60 (95)	50 (98)
Cancer type, n (%)			
Breast	44 (38.6)	22 (35)	22 (43)
Gynecological	13 (11.4)	6 (10)	7 (14)
Other	36 (31.6)	21 (33)	15 (29)
Not reported	21 (18.4)	14 (22)	7 (14)

[a]n=75 because of missing data.

[b]n=27 because of missing data.

[c]n=48 because of missing data.

Implementation

A total of 114 people were recruited to the program, of which 102 participants completed 3 or more sessions (89.4% completion of the program). The mean number of sessions was 5.0 (SD 1.5), and 61.4% (70/114) of participants attended all 6 sessions. Half of those who completed 3 or more sessions of the program (n=51) went on to complete the outcome measures, giving a 50% response rate to questionnaires at the end of the program.

Acceptability and Practicality

The 59 participant responses to the satisfaction questionnaire are detailed in Multimedia Appendix 2. The overwhelming majority of participants *strongly agreed* or *agreed* that the program and its delivery were positive and activities in the program were useful.

Preliminary Efficacy Testing

Table 3 provides a summary of all outcome measures for the 51 participants who completed baseline and postprogram questionnaires.

Table 3. Baseline and postprogram outcome scores (n=51).

Outcome measure[a,b]	Baseline scores, mean (SD)	Baseline scores, median (range)	Postprogram scores, mean (SD)	Postprogram scores, median (range)	*P* value[c]	*r* value[d]
Depression	8.9 (5.2)	9 (0-20)	6.1 (4.9)	5 (0-25)	<.001	−0.46
Anxiety	6.8 (4.9)	5 (0-20)	4.1 (4.1)	3 (0-18)	<.001	−0.35
Positive mental well-being	42.2 (9.3)	43 (21-63)	49.9 (7.5)	51 (25-67)	<.001	−0.50
Fatigue	18.1 (5.7)	19 (5-28)	14.4 (6.4)	13 (4-27)	<.001	−0.44
Fear of recurrence	15.9 (6.3)	13 (8-28)	13.2 (6.0)	12 (4-28)	<.001	−0.37
Hope	28.0 (10.0)	28 (8-47)	35.3 (7.4)	36 (11-46)	<.001	−0.51
Gratitude	34.8 (5.3)	36 (16-42)	36.7 (5.1)	38 (14-42)	.02	−0.24

[a]For depression, anxiety, fatigue, and fear of recurrence measures, decreasing scores indicate improvement.

[b]For positive mental well-being, hope, and gratitude measures, increasing scores indicate improvement.

[c]*P* value based on Wilcoxon signed-ranks.

[d]Effect size of change.

Psychological Distress

At the end of the iHOPE program, participants reported statistically significant improvements in depression and anxiety, with moderate effect sizes. Before the program, 47% (24/51) of participants exceeded the clinical cutoff scores for depression, and 43% (22/51) of participants exceeded the clinical cutoff scores for anxiety. At the end of the program, 31% (16/51) and 29% (15/51) of participants had scores that indicated a *recovery* from depression (z=−4.607; P<.001; r=−0.46) and anxiety (z=−3.534; P<.001; r=−0.35), respectively (ie, their scores were no longer above the threshold for clinical levels of distress).

Quality of Life for Cancer-Related Concerns

At the end of the iHOPE program, participants reported statistically significant improvements in cancer-related fatigue (z=−4.421; P<.001; r=−0.44) and *worry* or fear of recurrence (z=−3.765; P<.001; r=−0.37), with moderate effect sizes.

Positive Well-Being, Gratitude, and Hope

At the end of the iHOPE program, participants reported statistically significant improvements in positive mental well-being (z=−5.075; P<.001; r=−0.50) and hope (z=−5.113; P<.001; r=−0.51), with a moderate effect size for well-being and a large effect size for hope. Participants also reported statistically significant improvements in gratitude (z=−2.422; P=.015; r=−0.24), but with a small effect size (Table 3).

Discussion

Principal Findings

This study aimed to describe the development, evaluation, and feasibility of a digital self-management program for people who are living with, or have survived, all types of cancer. The program was well attended, with 114 participants at enrolment and 102 participants (89.5%) completing at least three of the six sessions. Feedback from the participants who completed the satisfaction questionnaire was overwhelmingly positive for many aspects of the content and design. Preliminary efficacy testing had positive results, with generally moderate-sized effects in the expected direction.

The majority of the participants were female, white, married, employed, and educated. Although this may present a limitation in terms of generalizing the study results to other demographic groups (eg, genders, cancer sites, ethnicities, and socioeconomic groups), a recent systematic review has reported that the majority of participants in self-management cancer programs are indeed women [26], and so, this study is representative in this respect. A low attendance rate for men is common in self-management and is linked to their reluctance to seek help [70]. Men are more likely to respond to marketing and recruitment messages that emphasize stoicism, independence, and control [70] and where marketing and recruitment materials contain images of men [71]. Our research group has begun looking at the self-management priorities of men diagnosed with cancer and the types of digital support they prefer [72,73]. It will be vital to conduct further development work to explore how to reach males, members of other ethnic groups, unemployed, and people with lower educational attainment. Ensuring that recruitment materials contain images and messages that appeal to multiple audiences and that recruitment and advertising take place in areas and locations frequented by people of all ages, ethnicities, genders, and income groups has been shown to widen participation in other studies [71]. These recruitment strategies will be suggested to MCS for future recruitment to the digital iHOPE program, as recommendations to reduce inequalities in care provision and recruitment bias.

Reporting rates of cancer sites were suboptimal, and data collection procedures should be improved, perhaps with a user-friendly checklist of cancer sites. Furthermore, collection of data on time since diagnosis and/or stage of treatment will allow us to verify in future studies if those recently diagnosed with cancer will show the same postprogram improvements in outcome measures as those who have survived cancer (or are coming to the end of treatment). Of those providing data, the largest group was breast cancer, reflecting national data. Gynecological cancers were reported by 11.4% of participants, which is greater than the estimated 5% of the population incidence [74]. However, the iHOPE program recognizes commonalities across types of cancer in the challenges faced and in unmet psychosocial needs; therefore, the majority of the

content is relevant to all cancer survivors regardless of the type of cancer. This pragmatic approach is a particular strength of the iHOPE program, especially as some patients may have cancer in more than one primary cancer site. Nevertheless, there are some elements of the iHOPE program content that may be more or less applicable to certain groups. In future cohorts, the iHOPE program could be offered to patients at all stages of treatment and/or with the same type of cancer. Another option is to adapt the program for some of the most common types of cancers, so that depending on the users' profile, personalized content is delivered. For example, in the body image session, prostate cancer and breast cancer users are provided with information, case studies, and other materials specific to their needs. Research has shown that personalized content improves engagement and retention.

The implementation of the program had positive results—89% of the participants completed the program (ie, attended 3 or more sessions) [51], and the average number of sessions attended was 5. Overall, 61% completed all 6 sessions, which is in line with rates of full program completion in other digital interventions, where a median of 56% of participants completed the full program [51,75].

There is a distinction between *dropout attrition* where participants do not fill in questionnaires (ie, 50% in this study) and *nonusage attrition* where participants stopped using the program (ie, only 11% completed <3 sessions in this study) [76]. High rates of nonusage attrition [76], are common and of concern in digitally delivered interventions. Reporting and analyzing the data at the level of session completion is crude, and a more useful and nuanced understanding of engagement and attrition is being developed, but still in its infancy. Researchers are beginning to use a variety of tools such as visualization and log data analysis, which provide evidence of features and content usage over time and detect usability issues. This type of analysis has the potential to improve user experience and improve engagement and outcomes. It is worth noting that some studies have not shown a linear relationship between time spent, the number of sessions completed, and outcomes [51]. It is possible that the *nonusage* participants in this study may be *e-attainers* (as described in [77]), where these participants may have left the program before completion but achieved what they needed from the program, such as learning about goal setting and stress management or obtaining reassurance and relief that their challenges and concerns are shared by others [37]. A higher rate of noncompletion of postprogram measures was observed in this digital delivery of iHOPE (ie, 50%) than the response rate in a face-to-face trial of the cancer HOPE program (88%, unpublished data). We suggest that the high noncompletion rate in this study was partly because of the timing of the research—data collection occurred around the Christmas period. Avoiding holiday times [78], using behavioral prompts [79], and reducing the number of outcome measures to reduce respondent burden [80] could improve data completion rates in future studies assessing the feasibility and practicality of digital cancer self-management programs.

The program was acceptable and appears to be practical, with positive feedback on program activities. The program uses peer support features and is supported by a facilitator with personal experience of cancer. There is a growing evidence base showing that trained peers can respond safely and therapeutically to distressing issues that often arise during self-management programs [81]. This study suggests the potential for the web environment to also provide peer support benefits to cancer survivors. The outcomes we measured capture important challenges for cancer survivors. Several activities in the iHOPE program directly address fatigue and worry, with a fatigue management session comprising video material and pacing and fatigue diaries, in addition to sessions on relaxation and mindfulness [82].

Acknowledging the significant limitations owing to the 50% completion rate of postprogram questionnaires, the findings show that scores for depression, anxiety, well-being, cancer-related fatigue and worry, hope, and gratitude were improved postprogram for those who completed the questionnaires. Moderate effect size improvements were achieved for most of the outcomes, which is consistent with other self-management research for long-term conditions involving self-selecting participants [83]. The fact that some of these improvements were evident in such a short period (6 weeks for postprogram outcomes) is extremely encouraging and consistent with early improvements found in brief self-management programs [84]. The variables likely to be used as clinical outcome variables (depression, anxiety, well-being, fatigue, and worry) all achieved moderate effect sizes in our uncontrolled pre-post study. Considering a future randomized controlled trial and taking into account the limitations of this study, a generic sample size calculation for moderate effect size (power 0.95, comparing 2 groups over 2 time points) would lead to a required total sample size of 132 (66 in each arm) [85].

The need for better care for comorbid depression in cancer survivors has been called for [14,19]. Many participants exceeded clinical cutoff values for depression (47%) and anxiety (43%) when they started the program. This indicates a high rate of clinical depression and anxiety among this population, therefore supporting the need for, and provision of, programs such as iHOPE. Further resources are needed to provide a range of interventions across the NHS for people with depression and anxiety following cancer. These resources could include a digital self-management program as part of a comprehensive stepped care package, particularly as the provision of expert mental health care across Europe remains limited [86].

By the end of the iHOPE program, the scores indicated that nearly one-third (16/51, 31%) had *recovered* from depression and over a quarter (15/51, 29%) had *recovered* from anxiety. These results are encouraging, compare well with general data on short-term psychological interventions [87], and suggest that further research investigating the effectiveness of the iHOPE program is imperative. The improved scores in gratitude and hope are reassuring, as this program is based on these core positive psychology concepts. Gratitude is linked to fewer health complaints, feeling more attentive, more energetic, more determined, more satisfied with life, more optimistic, having more feelings of connection to others, and being more likely to give and use social support [88]. The iHOPE program aims to increase hope among participants by encouraging participants to set and achieve weekly goals. Weekly goal setting and

feedback are important factors in initiating behavior change [89]. To date, there are very few hope-based, goal setting interventions that have been evaluated among cancer survivors [56], supporting the need for further development of the iHOPE program. An important next step is to formally explore if changes in hope and gratitude are important mechanisms of change for depression, anxiety, quality of life, and positive mental well-being.

Limitations

Our findings from this study are limited by the lack of a control group; lack of longer-term follow-up; small sample size; and a sample mainly of white, married women with a high level of education. Participants were recruited via the MCS charity website, Facebook, and other social media sites. This is likely to have attracted highly motivated cancer survivors, with good health and digital literacy skills, compared with cancer survivors who do not access health websites. Further studies of the feasibility and acceptability of the iHOPE program could utilize recruitment strategies aimed at widening participation for male cancer survivors [72,73] and participants from ethnic minorities and low-income backgrounds [71]. Further research is required to understand what drives some participants toward a digital intervention and to work with those with cancers in other sites to ensure that there are no barriers to accessing the iHOPE program.

There was a low response to the postprogram questionnaire. It is possible that there is a bias in responding to those who did find the program useful. The results of this study should be considered with caution, particularly in relation to the generalization of data. However, this feasibility testing suggests there is acceptability and practicality to our program, but there is a further requirement to support retention to complete research measures in future programs. This study used a crude usage attrition assessment. Future studies would benefit from using more sophisticated tracking tools and data analysis to reduce usage attrition and improve engagement and outcomes. Future research should now be a randomized controlled trial, powered to detect significant changes, a longer follow-up, and continued measurement of health and well-being related outcomes, satisfaction, and program usage.

Conclusions

We have shown that a digital self-management program, iHOPE, has the potential to improve several common and pressing unmet needs of cancer survivors, including fatigue and worry, depression, and anxiety. These improvements need to be further tested using a more robust research design involving a much larger sample over a longer time frame within a randomized controlled trial.

Acknowledgments

The authors would like to thank the individuals who took part in the initial usability testing of the iHOPE program and all those who participated in the study and completed the outcome measures. The authors would also like to thank the facilitators, Ian Cockerham, Lisa Larkin, Vicky Harker, and Stephen Wilkinson, who moderated the iHOPE program, and Macmillan Cancer Support, who funded the study.

Conflicts of Interest

AT is the coinventor of the HOPE program and iHOPE. WC, BW, GP, FM, LM, HW, and MM declare that they have no conflicts of interest.

Multimedia Appendix 1
Participant responses to the satisfaction and usability questionnaire.
[PNG File , 60 KB - jmir_v22i5e17824_app1.png]

Multimedia Appendix 2
Baseline scores for outcome variables between those who did (n=51) and did not (n=63) complete post-programme questionnaires.
[PNG File , 26 KB - jmir_v22i5e17824_app2.png]

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018 Nov;68(6):394-424 [FREE Full text] [doi: 10.3322/caac.21492] [Medline: 30207593]
2. Bray F, Jemal A, Grey N, Ferlay J, Forman D. Global cancer transitions according to the human development index (2008-2030): a population-based study. Lancet Oncol 2012 Aug;13(8):790-801. [doi: 10.1016/S1470-2045(12)70211-5] [Medline: 22658655]
3. Stewart B, Wild CP. World Cancer Report 2014. Geneva, Switzerland: World Health Organization; 2017.
4. Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, et al. Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer; 2018.

5. Maddams J, Utley M, Møller H. Projections of cancer prevalence in the United Kingdom, 2010-2040. Br J Cancer 2012 Sep 25;107(7):1195-1202 [FREE Full text] [doi: 10.1038/bjc.2012.366] [Medline: 22892390]
6. Armes J, Crowe M, Colbourne L, Morgan H, Murrells T, Oakley C, et al. Patients' supportive care needs beyond the end of cancer treatment: a prospective, longitudinal survey. J Clin Oncol 2009 Dec 20;27(36):6172-6179. [doi: 10.1200/JCO.2009.22.5151] [Medline: 19884548]
7. Foster C, Fenlon D. Recovery and self-management support following primary cancer treatment. Br J Cancer 2011 Nov 8;105(Suppl 1):S21-S28 [FREE Full text] [doi: 10.1038/bjc.2011.419] [Medline: 22048029]
8. Elliott J, Fallows A, Staetsky L, Smith PW, Foster CL, Maher EJ, et al. The health and well-being of cancer survivors in the UK: findings from a population-based survey. Br J Cancer 2011 Nov 8;105(Suppl 1):S11-S20 [FREE Full text] [doi: 10.1038/bjc.2011.418] [Medline: 22048028]
9. Cella D, Davis K, Breitbart W, Curt G, Fatigue Coalition. Cancer-related fatigue: prevalence of proposed diagnostic criteria in a United States sample of cancer survivors. J Clin Oncol 2001 Jul 15;19(14):3385-3391. [doi: 10.1200/JCO.2001.19.14.3385] [Medline: 11454886]
10. Hofman M, Ryan JL, Figueroa-Moseley CD, Jean-Pierre P, Morrow GR. Cancer-related fatigue: the scale of the problem. Oncologist 2007;12(Suppl 1):4-10 [FREE Full text] [doi: 10.1634/theoncologist.12-S1-4] [Medline: 17573451]
11. Simard S, Thewes B, Humphris G, Dixon M, Hayden C, Mireskandari S, et al. Fear of cancer recurrence in adult cancer survivors: a systematic review of quantitative studies. J Cancer Surviv 2013 Sep;7(3):300-322. [doi: 10.1007/s11764-013-0272-z] [Medline: 23475398]
12. Simard S, Savard J, Ivers H. Fear of cancer recurrence: specific profiles and nature of intrusive thoughts. J Cancer Surviv 2010 Dec;4(4):361-371. [doi: 10.1007/s11764-010-0136-8] [Medline: 20617394]
13. Koch L, Bertram H, Eberle A, Holleczek B, Schmid-Höpfner S, Waldmann A, et al. Fear of recurrence in long-term breast cancer survivors-still an issue. Results on prevalence, determinants, and the association with quality of life and depression from the cancer survivorship-a multi-regional population-based study. Psychooncology 2014 May;23(5):547-554. [doi: 10.1002/pon.3452] [Medline: 24293081]
14. Koch L, Jansen L, Brenner H, Arndt V. Fear of recurrence and disease progression in long-term (≥ 5 years) cancer survivors-a systematic review of quantitative studies. Psychooncology 2013 Jan;22(1):1-11. [doi: 10.1002/pon.3022] [Medline: 22232030]
15. Mitchell AJ, Ferguson DW, Gill J, Paul J, Symonds P. Depression and anxiety in long-term cancer survivors compared with spouses and healthy controls: a systematic review and meta-analysis. Lancet Oncol 2013 Jul;14(8):721-732. [doi: 10.1016/S1470-2045(13)70244-4] [Medline: 23759376]
16. Inhestern L, Beierlein V, Bultmann JC, Möller B, Romer G, Koch U, et al. Anxiety and depression in working-age cancer survivors: a register-based study. BMC Cancer 2017 May 19;17(1):347 [FREE Full text] [doi: 10.1186/s12885-017-3347-9] [Medline: 28526007]
17. NHS England. 2014. Five Year Forward View URL: https://www.england.nhs.uk/wp-content/uploads/2014/10/5yfv-web.pdf [accessed 2019-07-01]
18. NHS England. 2019. The NHS Long Term Plan URL: https://www.longtermplan.nhs.uk/publication/nhs-long-term-plan/ [accessed 2019-11-01]
19. Government of UK. 2013. Living With and Beyond Cancer: Taking Action to Improve Outcomes URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/181054/9333-TSO-2900664-NCSI_Report_FINAL.pdf [accessed 2020-03-01]
20. Griffiths C, Foster G, Ramsay J, Eldridge S, Taylor S. How effective are expert patient (lay led) education programmes for chronic disease? Br Med J 2007 Jun 16;334(7606):1254-1256 [FREE Full text] [doi: 10.1136/bmj.39227.698785.47] [Medline: 17569933]
21. Kim AR, Park H. Web-based self-management support interventions for cancer survivors: a systematic review and meta-analyses. Stud Health Technol Inform 2015;216:142-147. [Medline: 26262027]
22. Boland L, Bennett K, Connolly D. Self-management interventions for cancer survivors: a systematic review. Support Care Cancer 2018 May;26(5):1585-1595. [doi: 10.1007/s00520-017-3999-7] [Medline: 29199362]
23. Howell D, Richardson A, May C, Calman L, Fazelzad R, Moradian S, et al. Implementation of self-management support in cancer care and normalization into routine practice: a systematic scoping literature review protocol. Syst Rev 2019 Jan 31;8(1):37 [FREE Full text] [doi: 10.1186/s13643-019-0952-5] [Medline: 30704509]
24. Grimmett C, Armes J, Breckons M, Calman L, Corner J, Fenlon D, et al. RESTORE: an exploratory trial of an online intervention to enhance self-efficacy to manage problems associated with cancer-related fatigue following primary cancer treatment: study protocol for a randomized controlled trial. Trials 2013 Jun 21;14:184 [FREE Full text] [doi: 10.1186/1745-6215-14-184] [Medline: 23786716]
25. Frieden TR. Six components necessary for effective public health program implementation. Am J Public Health 2014 Jan;104(1):17-22. [doi: 10.2105/AJPH.2013.301608] [Medline: 24228653]
26. Cuthbert CA, Farragher JF, Hemmelgarn BR, Ding Q, McKinnon GP, Cheung WY. Self-management interventions for cancer survivors: a systematic review and evaluation of intervention content and theories. Psychooncology 2019 Nov;28(11):2119-2140. [doi: 10.1002/pon.5215] [Medline: 31475766]

27. Seligman ME. Learned Optimism: How to Change Your Mind and Your Life. New York City, USA: Knopf; 1991.
28. Seligman ME. Authentic Happiness: Using the New Positive Psychology to Realize Your Potential for Lasting Fulfillment. New York, USA: Free Press; 2002.
29. Linley PA, Joseph S, editors. Positive Psychology in Practice. New York, USA: John Wiley & Sons; 2004.
30. Tugade MM, Fredrickson BL, Barrett LF. Psychological resilience and positive emotional granularity: examining the benefits of positive emotions on coping and health. J Pers 2004 Dec;72(6):1161-1190 [FREE Full text] [doi: 10.1111/j.1467-6494.2004.00294.x] [Medline: 15509280]
31. Fredrickson BL. The role of positive emotions in positive psychology. The broaden-and-build theory of positive emotions. Am Psychol 2001 Mar;56(3):218-226 [FREE Full text] [doi: 10.1037//0003-066x.56.3.218] [Medline: 11315248]
32. Snyder CR, Harris C, Anderson JR, Holleran SA, Irving LM, Sigmon ST, et al. The will and the ways: development and validation of an individual-differences measure of hope. J Pers Soc Psychol 1991 Apr;60(4):570-585. [doi: 10.1037//0022-3514.60.4.570] [Medline: 2037968]
33. Snyder CR. Hypothesis: there is hope. In: Handbook of Hope: Theory, Measures, and Applications. New York, USA: Academic Press; 2000:3-18.
34. McCullough ME, Tsang J, Emmons RA. Gratitude in intermediate affective terrain: links of grateful moods to individual differences and daily emotional experience. J Pers Soc Psychol 2004 Feb;86(2):295-309. [doi: 10.1037/0022-3514.86.2.295] [Medline: 14769085]
35. Wood A, Maltby J, Gillett R, Linley PA, Joseph S. The role of gratitude in the development of social support, stress, and depression: two longitudinal studies. J Res Pers 2008 Aug;42(4):854-871 [FREE Full text] [doi: 10.1016/j.jrp.2007.11.003]
36. Bono G, McCullough ME. Positive responses to benefit and harm: bringing forgiveness and gratitude into cognitive psychotherapy. J Cogn Psychother 2006 Jun 1;20(2):147-158. [doi: 10.1891/jcop.20.2.147]
37. Yalom ID, Leszcz M. Theory and Practice of Group Psychotherapy. Fifth Edition. New York, USA: Basic Books; 2005.
38. Turner A. HOPE: a positive psychological group self-management support programme for cancer survivors. Psycho-Oncology 2012;21(S2):16. [doi: 10.1111/j.1099-1611.2012.03051.x]
39. Turner A. Systematic Development of a General Self-Management Intervention for Survivors of Cancer. In: Proceedings of the International Conference on Support for Self Management of Health. 2010 Presented at: SSMH'10; May 11-13, 2010; Stirling, UK.
40. Turner A, Batehup L, Surendranath S. Hope is key to recovery. Nurs Stand 2011;25(18):20-21. [doi: 10.7748/ns.25.18.20.s26] [Medline: 21309317]
41. Wilson PM. The UK expert patients program: lessons learned and implications for cancer survivors' self-care support programs. J Cancer Surviv 2008 Mar;2(1):45-52. [doi: 10.1007/s11764-007-0040-z] [Medline: 18648986]
42. Gerber BS. The chronic disease self-management program: extending reach through the internet. Med Care 2006 Nov;44(11):961-963. [doi: 10.1097/01.mlr.0000244923.80923.1e] [Medline: 17063126]
43. Lorig KR, Ritter PL, Laurent DD, Plant K. Internet-based chronic disease self-management: a randomized trial. Med Care 2006 Nov;44(11):964-971. [doi: 10.1097/01.mlr.0000233678.80203.c1] [Medline: 17063127]
44. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: a systematic review. J Med Internet Res 2016 May 16;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
45. Silva EH, Lawler S, Langbecker D. The effectiveness of mhealth for self-management in improving pain, psychological distress, fatigue, and sleep in cancer survivors: a systematic review. J Cancer Surviv 2019 Feb;13(1):97-107. [doi: 10.1007/s11764-018-0730-8] [Medline: 30635865]
46. Baumann LC, Dang TT. Helping patients with chronic conditions overcome barriers to self-care. Nurse Pract 2012 Mar 13;37(3):32-8; quiz 38. [doi: 10.1097/01.NPR.0000411104.12617.64] [Medline: 22289883]
47. Moody L, Turner A, Osmond J, Hooker L, Kosmala-Anderson J, Batehup L. Web-based self-management for young cancer survivors: consideration of user requirements and barriers to implementation. J Cancer Surviv 2015 Jun;9(2):188-200 [FREE Full text] [doi: 10.1007/s11764-014-0400-4] [Medline: 25236685]
48. 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 Jul 16;10:67 [FREE Full text] [doi: 10.1186/1471-2288-10-67] [Medline: 20637084]
49. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, PAFS Consensus Group. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. Pilot Feasibility Stud 2016;2:64 [FREE Full text] [doi: 10.1186/s40814-016-0105-8] [Medline: 27965879]
50. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. Am J Prev Med 2009 May;36(5):452-457 [FREE Full text] [doi: 10.1016/j.amepre.2009.02.002] [Medline: 19362699]
51. Enrique A, Palacios JE, Ryan H, Richards D. Exploring the relationship between usage and outcomes of an internet-based intervention for individuals with depressive symptoms: secondary analysis of data from a randomized controlled trial. J Med Internet Res 2019 Aug 1;21(8):e12775 [FREE Full text] [doi: 10.2196/12775] [Medline: 31373272]

52. Martin F, Turner A, Bourne C, Batehup L. Development and qualitative evaluation of a self-management workshop for testicular cancer survivor-initiated follow-up. Oncol Nurs Forum 2013 Jan;40(1):E14-E23. [doi: 10.1188/13.ONF.E14-E23] [Medline: 23269777]
53. Turner A, Martin F. Spreading HOPE: the development of a hope-based self-management intervention. In: Martz E, editor. Promoting Self-Management of Chronic Health Conditions: Theories and Practice. Oxford, UK: Oxford University Press; 2017:58-79.
54. Pearce G, Parke HL, Pinnock H, Epiphaniou E, Bourne CL, Sheikh A, et al. The PRISMS taxonomy of self-management support: derivation of a novel taxonomy and initial testing of its utility. J Health Serv Res Policy 2016 Apr;21(2):73-82. [doi: 10.1177/1355819615602725] [Medline: 26377727]
55. Abras C, Maloney-Krichmar D, Preece J. User-centered design. In: Bainbridge W, editor. Encyclopedia of Human-Computer Interaction. Thousand Oaks, CA: Sage Publications; 2004:445-456.
56. Casellas-Grau A, Font A, Vives J. Positive psychology interventions in breast cancer. A systematic review. Psychooncology 2014 Jan;23(1):9-19. [doi: 10.1002/pon.3353] [Medline: 23897834]
57. Lozano LM, García-Cueto E, Muñiz J. Effect of the number of response categories on the reliability and validity of rating scales. Methodology 2008 Jan;4(2):73-79. [doi: 10.1027/1614-2241.4.2.73]
58. Stinson J, McGrath P, Hodnett E, Feldman B, Duffy C, Huber A, et al. Usability testing of an online self-management program for adolescents with juvenile idiopathic arthritis. J Med Internet Res 2010 Jul 29;12(3):e30 [FREE Full text] [doi: 10.2196/jmir.1349] [Medline: 20675293]
59. Ammerlaan JW, Mulder OK, de Boer-Nijhof NC, Maat B, Kruize AA, van Laar J, et al. Building a tailored, patient-guided, web-based self-management intervention 'ReumaUitgedaagd!' for adults with a rheumatic disease: results of a usability study and design for a randomized control trail. JMIR Res Protoc 2016 Jun 23;5(2):e113 [FREE Full text] [doi: 10.2196/resprot.5735] [Medline: 27339472]
60. Fu MR, Axelrod D, Guth AA, Wang Y, Scagliola J, Hiotis K, et al. Usability and feasibility of health IT interventions to enhance self-care for Lymphedema symptom management in breast cancer survivors. Internet Interv 2016 Sep;5:56-64 [FREE Full text] [doi: 10.1016/j.invent.2016.08.001] [Medline: 28255542]
61. Avis NE, Smith KW, McGraw S, Smith RG, Petronis VM, Carver CS. Assessing quality of life in adult cancer survivors (QLACS). Qual Life Res 2005 May;14(4):1007-1023. [doi: 10.1007/s11136-004-2147-2] [Medline: 16041897]
62. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
63. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
64. Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, et al. The Warwick-Edinburgh mental well-being scale (WEMWBS): development and UK validation. Health Qual Life Outcomes 2007 Nov 27;5:63 [FREE Full text] [doi: 10.1186/1477-7525-5-63] [Medline: 18042300]
65. Maheswaran H, Weich S, Powell J, Stewart-Brown S. Evaluating the responsiveness of the Warwick Edinburgh mental well-being scale (WEMWBS): group and individual level analysis. Health Qual Life Outcomes 2012 Dec 27;10:156 [FREE Full text] [doi: 10.1186/1477-7525-10-156] [Medline: 23270465]
66. Snyder CR. Conceptualizing, measuring, and nurturing hope. J Couns Dev 1995;73(3):355-360. [doi: 10.1002/j.1556-6676.1995.tb01764.x]
67. Mccullough ME, Emmons RA, Tsang J. The grateful disposition: a conceptual and empirical topography. J Pers Soc Psychol 2002 Jan;82(1):112-127. [doi: 10.1037//0022-3514.82.1.112] [Medline: 11811629]
68. Fritz CO, Morris PE, Richler JJ. Effect size estimates: current use, calculations, and interpretation. J Exp Psychol Gen 2012 Feb;141(1):2-18. [doi: 10.1037/a0024338] [Medline: 21823805]
69. Rosnow RL, Rosenthal R. Definition and interpretation of interaction effects. Psychol Bull 1989;105(1):143-146. [doi: 10.1037/0033-2909.105.1.143]
70. Galdas P, Fell J, Bower P, Kidd L, Blickem C, McPherson K, et al. The effectiveness of self-management support interventions for men with long-term conditions: a systematic review and meta-analysis. BMJ Open 2015 Mar 20;5(3):e006620 [FREE Full text] [doi: 10.1136/bmjopen-2014-006620] [Medline: 25795688]
71. Horrell LN, Kneipp SM, Ahn S, Towne SD, Mingo CA, Ory MG, et al. Chronic disease self-management education courses: utilization by low-income, middle-aged participants. Int J Equity Health 2017 Jun 27;16(1):114 [FREE Full text] [doi: 10.1186/s12939-017-0604-0] [Medline: 28655319]
72. Batehup L, Cranshaw G, Lowson E, Lynall A, Martin F, Simmonds P, et al. Improving patient experience of cancer follow up: redesign and evaluation of adult cancer aftercare services for breast, colorectal, and testicular patients, at University Hospital Southampton NHS Foundation Trust. In: Proceedings of the National Cancer Survivorship Initiative Conference. 2012 Presented at: NCSI'12; November 16, 2012; London, UK.
73. Lynall A, Turner A, Szczepura A, Whiteman B. Identifying the factors that affect men's quality of life while waiting for confirmation of cancer diagnosis: a qualitative study. Eur J Surg Oncol 2018 Mar;44:S36 [FREE Full text] [doi: 10.1016/j.ejso.2018.01.565]

74. Cancer Research UK. 2019. Cancer Incidence for Common Cancers URL: https://www.cancerresearchuk.org/health-professional/cancer-statistics/incidence/common-cancers-compared#ref- [accessed 2019-07-01]
75. Waller R, Gilbody S. Barriers to the uptake of computerized cognitive behavioural therapy: a systematic review of the quantitative and qualitative evidence. Psychol Med 2009 May;39(5):705-712. [doi: 10.1017/S0033291708004224] [Medline: 18812006]
76. 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]
77. Martinez M. The E-Learning Guild. 2003. High Attrition Rates in e-Learning: Challenges, Predictors, and Solutions URL: https://www.elearningguild.com/pdf/2/071403mgt-l.pdf [accessed 2020-04-27]
78. Finkelstein SM, Celebrezze M, Cady R, Lunos S, Looman WS. Strategies to maximize data collection response rates in a randomized control trial focused on children with medical complexity. Telemed J E Health 2016 Apr;22(4):295-301 [FREE Full text] [doi: 10.1089/tmj.2015.0069] [Medline: 26484851]
79. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of prompts to promote engagement with digital interventions: a systematic review. J Med Internet Res 2016 Jan 8;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]
80. Rolstad S, Adler J, Rydén A. Response burden and questionnaire length: is shorter better? A review and meta-analysis. Value Health 2011 Dec;14(8):1101-1108 [FREE Full text] [doi: 10.1016/j.jval.2011.06.003] [Medline: 22152180]
81. Barlow JH, Bancroft GV, Turner AP. Self-management training for people with chronic disease: a shared learning experience. J Health Psychol 2005 Nov;10(6):863-872. [doi: 10.1177/1359105305057320] [Medline: 16176963]
82. Rouleau CR, Garland SN, Carlson LE. The impact of mindfulness-based interventions on symptom burden, positive psychological outcomes, and biomarkers in cancer patients. Cancer Manag Res 2015;7:121-131 [FREE Full text] [doi: 10.2147/CMAR.S64165] [Medline: 26064068]
83. Kennedy A, Reeves D, Bower P, Lee V, Middleton E, Richardson G, et al. The effectiveness and cost effectiveness of a national lay-led self care support programme for patients with long-term conditions: a pragmatic randomised controlled trial. J Epidemiol Community Health 2007 Mar;61(3):254-261 [FREE Full text] [doi: 10.1136/jech.2006.053538] [Medline: 17325405]
84. de Silva D. The Health Foundation. 2011. Evidence: Helping People Help Themselves URL: https://www.health.org.uk/publications/evidence-helping-people-help-themselves [accessed 2020-04-27]
85. Faul F. G*Power: Statistical Power Analyses for Windows and Mac. 2009. URL: http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/download-and-register [accessed 2020-04-27]
86. Travado L, Reis JC, Watson M, Borràs J. Psychosocial oncology care resources in Europe: a study under the European partnership for action against cancer (EPAAC). Psychooncology 2017 Apr;26(4):523-530. [doi: 10.1002/pon.4044] [Medline: 26685870]
87. Hofmann SG, Asnaani A, Vonk IJ, Sawyer AT, Fang A. The efficacy of cognitive behavioral therapy: a review of meta-analyses. Cognit Ther Res 2012 Oct 1;36(5):427-440 [FREE Full text] [doi: 10.1007/s10608-012-9476-1] [Medline: 23459093]
88. Emmons RA, McCullough ME. Counting blessings versus burdens: an experimental investigation of gratitude and subjective well-being in daily life. J Pers Soc Psychol 2003 Feb;84(2):377-389. [doi: 10.1037//0022-3514.84.2.377] [Medline: 12585811]
89. Hardeman W, Johnston M, Johnston D, Bonetti D, Wareham N, Kinmonth A. Application of the theory of planned behaviour in behaviour change interventions: a systematic review. Psychol Health 2002 Jan;17(2):123-158. [doi: 10.1080/08870440290013644a]

Abbreviations

HOPE: Help to Overcome Problems Effectively
iHOPE: Help to Overcome Problems Effectively (where ‘i’ indicates digital version)
MCS: Macmillan Cancer Support
NHS: National Health Service
QLACS: Quality of Life in Adult Cancer Survivors scale

Edited by G Eysenbach; submitted 15.01.20; peer-reviewed by J Thompson, H Goodfellow, R Whitney; comments to author 05.02.20; revised version received 19.03.20; accepted 21.03.20; published 19.05.20.

Please cite as:
Martin F, Wright H, Moody L, Whiteman B, McGillion M, Clyne W, Pearce G, Turner A
Help to Overcome Problems Effectively for Cancer Survivors: Development and Evaluation of a Digital Self-Management Program
J Med Internet Res 2020;22(5):e17824
URL: http://www.jmir.org/2020/5/e17824/
doi:10.2196/17824
PMID:32209529

Original Paper

Efficacy of Text Messaging and Personal Consultation by Pharmacy Students Among Adults With Hypertension: Randomized Controlled Trial

Panpan Zhai[1,2,3], BS; Khezar Hayat[1,2,3,4], MS; Wenjing Ji[1,2,3], PhD; Qian Li[1,2,3], BS; Li Shi[1,2,3], BS; Naveel Atif[1,2,3], MS; Sen Xu[1,2,3], BS; Pengchao Li[1,2,3], BS; Qianqian Du[1,2,3], BS; Yu Fang[1,2,3], PhD

[1]Department of Pharmacy Administration and Clinical Pharmacy, Xi'an Jiaotong University, Xi'an, China

[2]Center for Drug Safety and Policy Research, Xi'an Jiaotong University, Xi'an, China

[3]Shaanxi Center for Health Reform and Development Research, Xi'an Jiaotong University, Xi'an, China

[4]Institute of Pharmaceutical Sciences, University of Veterinary and Animal Sciences, Lahore, Pakistan

Corresponding Author:
Yu Fang, PhD
Department of Pharmacy Administration and Clinical Pharmacy
Xi'an Jiaotong University
No 76 Yanta West Road
Xi'an, 710061
China
Phone: 86 2982655132
Email: yufang@mail.xjtu.edu.cn

Abstract

Background: Hypertension is one of the leading risk factors for ischemic heart diseases, and high rates of hypertension prevalence have either remained the same or increased in developing countries in recent years. Unfortunately, about 20% to 50% of patients with chronic diseases have been nonadherent to their drug therapy. SMS text messaging and pharmacy student–led consultations have the potential to help patients manage their blood pressure (BP).

Objective: The aim of this study was to assess the effectiveness, feasibility, and acceptability of SMS text messaging and consultation to manage the BP of Chinese patients with hypertension.

Methods: We conducted a two-arm cluster randomized controlled trial among patients with hypertension in Xi'an City, Shaanxi Province, China, and recruited 384 patients from 8 community health care centers. Patients were randomized into an intervention group to receive SMS text messages and consultations or into a control group to receive usual care for 3 months. We sent SMS text messages at 3-day intervals and collected data at baseline (including demographics, clinical outcomes, medication complexity, side effects, patient behavior, knowledge about hypertension, BP, and medication adherence) and the 3-month follow-up (including BP, medication adherence, and knowledge about hypertension).

Results: We assessed 445 patients with hypertension and excluded 61 patients who were not eligible or who had not filled out their questionnaires. The mean age of the patients was 68.5 (SD 7.9) years in the intervention group and 69.4 (SD 9.7) years in the control group, and the sample was primarily female (265/384, 69.0%). Patients in the intervention group showed significant improvements in systolic BP (SBP; mean 134.5 mm Hg, SD 15.5 mm Hg vs mean 140.7 mm Hg, SD 15.2 mm Hg; *P*=.001), medication adherence (mean 7.4, SD 1.2 vs mean 7.0, SD 1.3; *P*=.04), and knowledge about hypertension (mean 6.3, SD 0.9 vs mean 5.9, SD 1.2; *P*=.004) compared with those in the control group. In measures of diastolic BP (DBP), the two arms showed nonsignificant improvements (mean 78.2 mm Hg, SD 9.0 mm Hg vs mean 77.2 mm Hg, SD 10.3 mm Hg; *P*=.06). In total, 176 patients had controlled BP at the 3-month follow-up (98 patients in the intervention group vs 78 patients in the control group), but it was nonsignificant (*P*=.08).

Conclusions: The use of SMS text messaging and consultation to manage SBP and improve medication adherence is effective, feasible, and acceptable among Chinese patients with hypertension, although a significant difference was not observed with regard to DBP. It is important to maximize the potential of SMS text messaging and consultation by increasing the feasibility and acceptance of mobile interventions and conduct a cost-effectiveness analysis on this method.

Trial Registration: Chinese Clinical Trial Registry ChiCTR1900026862; http://www.chictr.org.cn/showproj.aspx?proj=42717.

(J Med Internet Res 2020;22(5):e16019) doi:10.2196/16019

KEYWORDS

medication adherence; text messaging; hypertension; consultation; pharmacy students

Introduction

Background

Hypertension is regarded as one of the leading risk factors for ischemic heart diseases, affecting more than 1.4 billion people globally [1,2]. In the United States alone, over 100 million people are living with hypertension, according to a recent report of the American Heart Association [3]. Hypertension has also greatly amplified the rate of deaths, for example, every year nearly 9.4 million deaths are attributed to hypertension worldwide [4]. High rates of hypertension prevalence have either remained the same or increased in developing countries in recent years [5]. A study conducted in China from 2012 to 2015 concluded that the prevalence of hypertension in China was 27.9% [6], increased by 48.4% compared with the previous study conducted in 2002 [7].

The risk of ischemic heart diseases could be reduced by up to 50% via lowering the systolic blood pressure (SBP) by 20 mm Hg or diastolic blood pressure (DBP) by 10 mm Hg, thus preventing the rate of mortality and morbidity [8]. In the People's Republic of China, hypertension, with its several life-threatening complications, including stroke, heart attack, heart failure, renal failure, and coma, has seriously challenged the financial health care delivery system [9].

The management of hypertension is based on drug therapy coupled with lifestyle modifications. Better control of hypertension is possible by improving adherence to antihypertensive medications, which could further minimize its complications [10]. Unfortunately, adherence to the prescribed medications is always considered as a problem in patients with chronic diseases [11]. For example, a study reported that 20% to 50% of patients with chronic disorders have been nonadherent toward their drug therapy [12-14]. Similarly, patients with hypertension have also struggled to comply with the guidelines of their health care providers [15-17].

A large number of interventions and strategies have been proposed to improve adherence to the treatment plan among patients despite inconsistent results [18]. One such approach is the use of SMS text messaging via mobile technology, which has the potential to tailor patients' behavior to support their treatment adherence [19,20]. The use of customized SMS text messages offers several advantages, including cost-efficiency, wide coverage, ease in accessibility, and timely delivery of useful medical information [21-23].

Pharmacists are an integral part of the health care system, and their role in overcoming poor medication adherence is evident in the literature [24]. Owing to the limited number of qualified pharmacists in developing counties and the lack of proper education and counseling to patients [25,26], patients are struggling to cope with the problem of poor adherence. To address this issue, consultations by pharmacy students could be another alternative that is already in use in different health care settings of the United States and Kenya [27,28]. Meanwhile, as future pharmacists, pharmacy students can expand the scope of pharmacists' practices by holding patient consultations in their early stages. A study highlighted that 89% of patients had a better understanding of their medications when counseled by pharmacy students and residents in a discharge counseling program [28]. Furthermore, a 5-min pharmacy student–led counseling session offered to 198 participants greatly improved the rates of vaccination as well as reduced medication errors and costs in acute health care settings [23].

Objectives

A systematic review concluded that an intervention needs to reduce costs and expand outreach, and more studies from resource-limited settings, especially in Africa and Asia, are needed [29]. Moreover, a greater number of studies have been conducted to determine the effect of SMS text messaging on adherence among patients with heart diseases, mental diseases, and HIV. The literature is lacking in studies investigating the impact of SMS text message reminders sent by pharmacy students on patients with hypertension [30-33]. Thus, we conducted a two-arm cluster randomized controlled trial among patients with hypertension to assess the effectiveness, feasibility and acceptability of SMS text messaging and consultation to manage blood pressure (BP) among Chinese patients with hypertension.

Methods

Study Design

From October 2018 to May 2019, we completed the randomized controlled trial testing the efficacy of SMS text messages and personal consultation to improve adherence among patients with hypertension. The study was reviewed and approved by the Research Ethics Committee of Health Science Center, Xi'an Jiaotong University (Number: 2018531), and all patients provided written informed consent before enrollment. We conducted this study in 8 community health care centers (CHCs) in Xi'an City, Shaanxi Province, China. The full details of the trial design were published previously [34]; however, key elements are summarized subsequently. This study compared medication adherence and clinical outcomes, including SBP and DBP, between the intervention and control groups. The 8 CHCs, located in different geographic and socioeconomic areas, were randomized in a ratio of 1:1 to the intervention or control group using a random number generated by the computer. The study is reported according to the Consolidated Standards of Reporting Trials of Electronic and Mobile HEalth Applications and onLine TeleHealth V1.6.1 (Multimedia Appendix 1) [35].

Participants

To be eligible for enrollment, patients had to be adults (aged over 18 years) meeting the following criteria: diagnosed with hypertension and currently using antihypertensive medications,

BP <220/120 mm Hg at enrollment, and with health records at CHCs included in the study. Participants also had to have a mobile phone capable of receiving SMS text messages or access to a mobile phone and be able to read and complete informed consent. We excluded patients who had dementia, depression, and serious heart, lung, and kidney diseases and who were pregnant or in their lactation period.

Patient Recruitment

Participants were recruited during the process of physical examination or physician consultation. Potential patients were screened and approached by the clinic staff and a trained research assistant. Eligible patients in both groups completed a baseline survey comprising demographics, clinical outcomes, medication complexity, side effects, patient behavior, knowledge about hypertension, and medication adherence. Medication adherence was measured by using a validated 8-item Morisky Medication Adherence Scale (MMAS-8) [36]. The 8 items of MMAS-8 were designed to detect patients who are nonadherent to their medications and the reasons for such behavior, such as forgetfulness, insufficient knowledge, inconvenience, and side effects. The score of MMAS-8 ranges from 0 to 8, with a score of <6 indicating low adherence, 6 to <8 indicating medium adherence, and 8 indicating high adherence. The knowledge questionnaire included 7 questions, and the score ranges from 0 to 7, with a high score indicating a high knowledge level.

Intervention

The intervention program comprised 2 components. The first component comprised personal consultations by trained pharmacy students. The second component was SMS text messages sent at 3-day intervals. Patients randomized into the intervention group were given a personal consultation that lasted about 5 min to identify the reasons for medication nonadherence. For patients with poor memory for taking medication, we suggested setting an alarm on their phones as well as keeping medications near their beds or drinking glasses. Regarding patients with poor knowledge, we helped them understand the need to take antihypertensive medications throughout their lives to decrease BP fluctuation as well as the chances of serious complications, and we told them to modify their lifestyles, for example, reducing salt intake or stopping tobacco or alcohol consumption. Pharmacy students certified by Professor Morisky took personal consultations. Professor Morisky and the research group trained pharmacy students on July 20 and 21, 2018. The training comprised 3 parts: interactive discussion on measurements of medication adherence and their advantages and disadvantages, Morisky Training Test, and use of the Morisky Widget. Professor Morisky trained the students on how to use MMAS-8 appropriately to detect patients with low medication adherence and identify the reasons behind it so that we could offer advice to overcome medication nonadherence on an individual basis. The students were asked to finish a test assessment after the first day's training, and the students who failed the test were excluded from further training. In addition, we conducted a role-playing process to simulate the real scenario when faced with patients. Professor Morisky evaluated the process and offered general and individual feedback. Only those students who passed the test assessment and role-playing evaluation were given certification by Professor Morisky. The details of the training sessions were discussed in our earlier study [34].

After a personal consultation, we told patients to read the SMS text messages sent by our research group. SMS text messages were unidirectional and sent at 3-day intervals at 7 AM. The content of the SMS text messages was developed by researchers, a nurse coordinator, and a cardiologist with regard to helpful suggestions on the management of risk factors and motivation based on previous references [10,20,37]. SMS text messages were incorporated into 3 parts: knowledge about hypertension, lifestyle modifications and additional information on physical activity, dietary sodium reduction, normal BP, complications of hypertension, healthy diet, smoking cessation, weight reduction, and strategies to cope with emotions [34], and suggestions of measures to improve medication adherence. SMS text messages were sent to patients via an open website. During the 3-month program, patients in the intervention group received a total of 30 SMS text messages. The examples of SMS text messages are shown in Table 1.

Table 1. Text message contents.

Category classification	Example of content
Knowledge about hypertension	Please remember to take your antihypertensive medications every day. Hypertension is defined as systolic blood pressure ≥140 mm Hg or diastolic blood pressure ≥90 mm Hg. Hypertension can cause cardiovascular disease, stroke, and kidney failure. Please keep an eye on your blood pressure and seek medical assistance if you have any abnormalities
Lifestyle modifications	Please remember to take your antihypertensive medications every day. Please reduce salt intake and pay attention to the intake of hidden salt (sausage, canned food, soy sauce, and pickles). The maximum of daily salt intake is a beer bottle cap (about 6 grams). Please keep an eye on your blood pressure and seek medical assistance if you have any abnormalities
Measures to improve medication adherence	Please remember to take your antihypertensive medications every day. Put medications around your toothbrush or drinking glass to remind yourself of taking them daily. Please keep an eye on your blood pressure and seek medical assistance if you have any abnormalities

Control Group

Patients in the control group received a welcome SMS text message and an end-of-trial SMS text message, but they did not receive a personal consultation. Besides, both patients in the intervention group and control group were given standard pharmaceutical care according to the Guidelines for Good Pharmacy Practice (GPP) [38]. GPP require pharmacists to provide medications along with pharmacy practice, which is a patient-centered, outcomes-oriented service to cooperate with

physicians and nurses to promote health, prevent and manage disease, and assess, monitor, initiate, and modify the medication regimen. Finally, patients' health-related quality of life can be optimized.

Follow-Up Assessment

Follow-up assessments were performed in both study arms at month 3, based on the intention-to-treat principles. In this case, outcomes were evaluated in all randomized participants, including those in the intervention arm who did not read SMS text messages. The study assessment also evaluated medication adherence, BP, and patient knowledge about hypertension at month 3.

Outcomes

The primary outcomes were a mean change of BP, measured by using the OMRON BP cuff (HEM-7200, OMRON Corporation) by clinical staff and trained research coordinators, as well as medication adherence and knowledge about hypertension at baseline and month 3. Two BP readings were collected after the patients sat and rested for 5 min [39]. A third reading was taken if the difference between the previous two readings was ≥5 mm Hg, and the final BP reading was the average of the two closest BP readings [39]. The secondary outcome was whether patients had controlled BP, defined as SBP<140 mm Hg and DBP<90 mm Hg for patients without diabetes mellitus or chronic kidney disease, SBP<130 mm Hg and DBP<80 mm Hg for patients with diabetes mellitus or chronic kidney disease, and SBP<150 mm Hg and DBP<90 mm Hg for patients aged over 80 years.

Sample Size

We sought to enroll 384 patients to have 90% power to detect a 5-mm Hg difference in BP between the intervention and control groups, with an alpha of .05, a 15% attrition rate, or an SD of up to 14.1 mm Hg [40].

Statistical Analysis

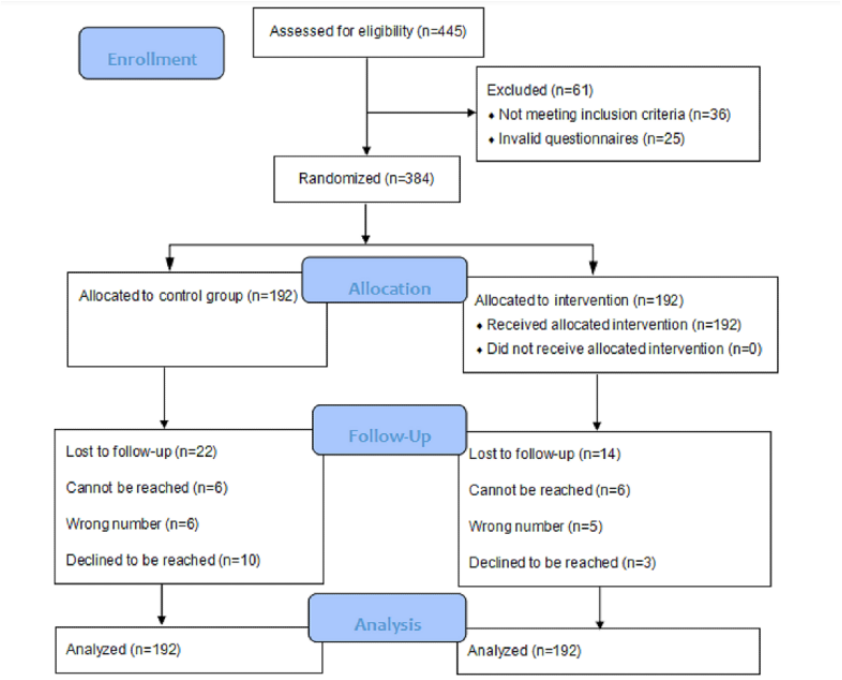
We conducted our analyses using intention-to-treat principles. Frequencies and percentages or means and SDs were presented in descriptive statistics. An independent-samples 2-tailed *t* test was used for continuous variables, and the chi-square test was used for categorical variables. A univariate linear regression model was used to analyze factors associated with medication adherence and SBP. The primary outcomes were analyzed using a mixed effect model to produce a difference in mean change, 95% CIs, and *P* values from baseline to follow-up. The model considered time, treatment group, and group-by-time interactions as fixed effects and incorporated random effects for individuals across time. The secondary outcome was analyzed using a mixed effect logistic model. All analyses were performed using SPSS Version 18.0 (SPSS Inc), and a two-sided *P* value of <.05 was considered statistically significant.

Results

Study Population

We assessed 445 patients with hypertension and excluded 61 patients who were not eligible or who had not filled their questionnaires. A total of 384 patients were randomly allocated to either the intervention group (n=192) or the control group (n=192). Figure 1 shows study enrollment and retention flow. A total of 14 and 22 patients were lost to follow-up in the intervention and control groups, respectively. In the final analysis, we included 384 patients in the mixed models based on the intention-to-treat analysis.

Figure 1. Flow diagram.



Baseline Characteristics

The mean age was 68.5 (SD 7.9) years in the intervention group and 69.4 (SD 9.7) years in the control group. The sample was primarily female (265/384, 69.0%), with 64.6% (124/192) of patients in the intervention group and 73.4% (141/192) in the control group. More than one-third (146/384, 38.0%) of the patients had attained junior high school education, and most of the patients did not smoke (343/384, 89.3%) or drink (340/384, 88.5%). A total of 83.1% (319/384) of the participants were living with family members; of them, 166 (52.0%) patients were in the intervention group, and 153 (48.0%) patients were in the control group. The mean BMI was 25.2 kg/m^2 and 24.8 kg/m^2 in the intervention group and control group, respectively. Their mean SBP and DBP were 147.9 mm Hg and 78.9 mm Hg, respectively, with a mean score of 5.6 in the knowledge part. The mean score of MMAS-8 was similar in both groups (mean 6.9, SD 1.5 in the intervention group and mean 6.8, SD 1.3 in the control group). Sociodemographic variables between the randomized groups were well balanced, and these are shown in Table 2.

Table 2. Baseline characteristics of participants (N=384).

Variables	Intervention group (n=192)	Control group (n=192)	*P* value
Age (years), mean (SD)	68.5 (7.9)	69.4 (9.7)	.32
Gender, n (%)			.08
Male	68 (35.4)	51 (26.6)	
Female	124 (64.6)	141 (73.4)	
Education, n (%)			.02
Primary school or below	37 (19.3)	56 (29.2)	
Junior high school	82 (42.7)	64 (33.3)	
High school	52 (27.1)	46 (24.0)	
College	9 (4.7)	4 (2.1)	
Undergraduate university or above	12 (6.3)	22 (11.5)	
Smoking, n (%)			.74
Yes	22 (11.5)	19 (9.9)	
No	170 (88.5)	173 (90.1)	
Drinking, n (%)			.08
Yes	28 (14.6)	16 (8.3)	
No	164 (85.4)	176 (91.7)	
Living arrangements, n (%)			.10
Living alone	26 (13.5)	39 (20.3)	
Living with family members	166 (86.5)	153 (79.7)	
BMI (kg/m^2), mean (SD)	25.2 (3.0)	24.8 (3.1)	.20
Systolic blood pressure (mm Hg), mean (SD)	146.0 (18.1)	149.9 (21.7)	.05
Diastolic blood pressure (mm Hg), mean (SD)	77.9 (10.7)	79.9 (10.6)	.06
8-item Morisky Medication Adherence Scale score, mean (SD)	6.9 (1.5)	6.8 (1.3)	.67
Knowledge score, mean (SD)	5.7 (1.3)	5.6 (1.2)	.34

Factors Associated With Medication Adherence and Systolic Blood Pressure

The determinants of medication adherence and SBP are presented in Tables 3 and 4, respectively. Age and frequency of BP measurements were found to have significant associations with medication adherence by the linear regression model. For age, beta was 0.19 (95% CI 0.01 to 0.05; *P*=.001), and patients with older age tended to be less likely to be medication nonadherent. For the frequency of BP measurements, patients with a higher frequency of BP measurements were more likely to be medication adherent (beta=.19; 95% CI 0.05 to 0.18; *P*<.001). With a linear regression model, age, gender, education, and DBP were found to be significantly associated with SBP. For age, young patients were likely to have optimal SBP (beta=.28; 95% CI 0.41 to 0.87; *P*<.001). Patients who were female tended to have higher SBP than their male counterparts (beta=.12; 95% CI 0.42 to 9.88; *P*=.03). Those with a bachelor's degree or above (beta=−.12; 95% CI −3.90 to −0.33; *P*=.02) and low DBP (beta=.48; 95% CI 0.72 to 1.08; *P*<.001) were likely to have low SBP.

Table 3. Factors associated with medication adherence.

Variables	Beta (95% CI)	*P* value[a]
Age (years)	0.19 (0.01 to 0.05)	.001
Gender	0.02 (−0.32 to 0.44)	.75
Education	0.01 (−0.13 to 0.16)	.87
Smoking	0.02 (−0.46 to 0.62)	.77
Drinking	0.06 (−0.23 to 0.79)	.28
Living arrangements	0.05 (−0.21 to 0.56)	.38
BMI (kg/m^2)	0.05 (−0.02 to 0.07)	.33
Systolic blood pressure (mm Hg)	−0.10 (−0.02 to 0.001)	.10
Diastolic blood pressure (mm Hg)	0.05 (−0.01 to 0.02)	.50
Knowledge score	0.02 (−0.10 to 0.14)	.72
Frequency of blood pressure measurement	0.19 (0.05 to 0.18)	<.001
Duration of hypertension	−0.05 (−0.15 to 0.06)	.37
Family member of hypertension	0.03 (−0.22 to 0.39)	.58
Number of antihypertensive drugs used	−0.01 (−0.37 to 0.30)	.86
Number of complications	−0.01 (−0.24 to 0.21)	.90

[a]Linear regression model.

Table 4. Factors associated with systolic blood pressure.

Variables	Beta (95% CI)	*P* value[a]
Age (years)	0.28 (0.41 to 0.87)	<.001
Gender	0.12 (0.42 to 9.88)	.03
Education	−0.12 (−3.90 to −0.33)	.02
Smoking	−0.02 (−8.00 to 5.43)	.71
Drinking	−0.01 (−7.10 to 5.63)	.82
Living arrangements	0.06 (−1.42 to 8.18)	.17
BMI (kg/m^2)	−0.02 (−0.72 to 0.45)	.65
Diastolic blood pressure (mm Hg)	0.48 (0.72 to 1.08)	<.001
8-item Morisky Medication Adherence Scale score	−0.08 (−2.35 to 0.20)	.10
Knowledge score	0.04 (−0.89 to 2.16)	.41
Frequency of blood pressure measurement	−0.00 (−0.81 to 0.82)	.99
Duration of hypertension	0.03 (−0.97 to 1.68)	.55
Family member of hypertension	0.05 (−1.81 to 5.80)	.30
Number of antihypertensive drugs used	−0.01 (−4.41 to 3.91)	.91
Number of complications	−0.07 (−4.90 to 0.77)	.15

[a]Linear regression model.

Primary Outcomes and Secondary Outcome

Table 5 shows the primary outcomes measured at baseline and follow-up and changes in these outcomes between the intervention and control groups. The mean (95% CI and *P* value) difference change in SBP between the two groups was −6.86 mm Hg (−10.37 to −3.34 and *P*=.001). Both intervention and control groups had an improved MMAS-8 score (mean 7.4, SD 1.2 and mean 7.0, SD 1.3, respectively) and a knowledge score (mean 6.3, SD 0.9 and mean 5.9, SD 1.2, respectively) at follow-up, and intervention patients experienced significant mean improvements in these two outcomes compared with the control group (*P*=.04 and *P*=.004). However, negligible improvements in DBP were noted (mean 78.2 mm Hg, SD 9.0 mm Hg and mean 77.2 mm Hg, SD 10.3 mm Hg, respectively) with no significant difference found between the two groups

(*P*=.06). The proportion of participants achieving controlled BP was 55.7% (98/176) and 44.3% (78/176) in the intervention group and control group, respectively, at the follow-up (Table 6), and the improvements in BP control did not have a significant difference between the two arms (*P*=.08).

Table 5. Baseline and changes in systolic blood pressure, diastolic blood pressure, 8-item Morisky Medication Adherence Scale score, and knowledge score at each assessment point using the missing data method.

Variables	Intervention group, mean (SD)	Control group, mean (SD)	Difference in mean change (95% CI)	*P* value[a]
Systolic blood pressure				.001
Baseline	146.0 (18.1)	149.9 (21.7)	N/A[b]	
3 months	134.5 (15.5)	140.7 (15.2)	−6.86 (−10.37 to −3.34)	
Diastolic blood pressure				.06
Baseline	77.9 (10.7)	79.9 (10.6)	N/A	
3 months	78.2 (9.0)	77.2 (10.3)	−0.95 (−1.09 to 2.98)	
8-item Morisky Medication Adherence Scale score				.04
Baseline	6.9 (1.5)	6.8 (1.3)	N/A	
3 months	7.4 (1.2)	7.0 (1.3)	0.39 (0.12 to 0.65)	
Knowledge score				.004
Baseline	5.7 (1.3)	5.6 (1.2)	N/A	
3 months	6.3 (0.9)	5.9 (1.2)	0.44 (0.22 to 0.66)	

[a]Mixed effect model was used.

[b]N/A: not applicable.

Table 6. Secondary outcome at the 3-month follow-up.

Variable - Blood pressure control[a]	Intervention group, n (%)	Control group, n (%)
Baseline	70 (54.3)	59 (45.7)
3 months	98 (55.7)	78 (44.3)

[a]Mixed effect logistic regression model showed a *P* value of .08.

Text Acceptability

We sent 5760 SMS text messages to participants in the intervention group and 384 SMS text messages to those in the control group, and all SMS text messages were sent successfully. In the intervention group, 68.0% (121/178) of the patients reported reading the messages sent by our research group, and 32.0% (57/178) of the patients reported having not read the messages as they did not have the habit of reading messages (34/57, 60%) and they could not see messages clearly because of old age (18/57, 32%), and some of them declared that they were too busy to read messages (5/57, 9%). A total of 69.7% (124/178) of participants in the intervention group reported having improved satisfaction about their BP, whereas 53 (29.8%) and 1 (0.6%) of them reported having the same satisfaction and decreased satisfaction, respectively, compared with the baseline study at month 3. The details are presented in Table 7.

Table 7. Text acceptability after the 3-month intervention (n=178).

Variables	Value, n (%)
How many SMS text messages that were sent by our research group did you read?	
0	57 (32.0)
1-10	55 (30.9)
10-20	29 (16.3)
20-30	37 (20.8)
What is the reason for not reading SMS text messages?	
I do not have the habit of reading messages	34 (59.6)
I cannot see messages clearly because of old age	18 (31.6)
I am too busy to read messages	5 (8.8)
Has your satisfaction with blood pressure control changed after participating in this study?	
Yes, my satisfaction increased over the process	124 (69.7)
No, my satisfaction remained the same	53 (29.8)
Yes, my satisfaction decreased over the process	1 (0.6)

Discussion

Principal Findings

In this randomized controlled trial of SMS text messaging and consultation, we found significant improvements in both medication adherence and SBP over 3 months of follow-up in the intervention group compared with the control group in a general population of patients with hypertension. Using an inexpensive means of communication and consultation, there was a relative increase of 0.5 in the MMAS-8 score and a decrease of 11.5 mm Hg in SBP in the intervention group. The secondary outcome of the trial focusing on patients having controlled BP showed a promising trend toward improved BP control even though the difference was insignificant.

There was no evidence of differences in the intervention effectiveness in DBP; however, compared with DBP, elevated SBP is the main target of antihypertensive therapy. Epidemiological studies indicated that hypertension is undertreated, especially SBP, which should be the primary target of antihypertensive treatment [41].

During the study period, patients reported high acceptability of SMS text messaging and consultation. They said that it helped them to understand how to manage hypertension through medication treatment and lifestyle modifications in their daily lives. However, there were participants (32.0%) who did not read their messages. Studies showed that improved patient-provider communication could help patients commit to lifestyle adaptations [42]. In future studies, we can call patients every month to confirm whether they read the SMS text messages. If they did not, we can remind them and ask about their interests in the context of SMS text messaging. In this way, patients are more engaged and involved in the process, so they will likely be more committed to the study.

A large number of previous studies based on SMS text messaging and student consultation have shown a significant improvement in medication adherence [33,43,44]. A meta-analysis of 16 randomized trials involving 2742 patients with chronic diseases, such as asthma, cardiovascular diseases, and HIV, has shown twice the improvement in the odds of medication adherence [43]. This improvement of medication adherence translated into adherence rates from 50% to 67.8%, with an absolute increase of 17.8%, suggesting that SMS text messaging interventions have the potential to improve medication adherence in patients with chronic diseases [43]. Moreover, a study conducted in Santiago, Chile, among 314 patients with hypertension with <6 months of antihypertensive treatment, found a boost in adherence levels from 49.0% to 62.3% in the SMS text messages–led intervention group [44]. Abughosh et al [33] evaluated the impact of an intervention based on motivational interviewing conducted by 11 pharmacy students on 743 patients with diabetes and hypertension. This intervention also resulted in a significant improvement in the patient's adherence to medication treatment.

Strengths and Limitations

However, there are a few limitations inherent in this study. First, we hypothesized that medication adherence and SBP have a significant association, but the hypothesized interaction was not found in the baseline survey. This finding contrasts with studies that have observed better BP control in those with better adherence [45]. Medication adherence was measured by a self-reported measure, which has social desirability bias and may overestimate true adherence [46]. Furthermore, patients were recruited with the help of nurses and physicians; therefore, patients may have been more likely to report being adherent to please the medical staff. Moreover, those with high baseline SBP and DBP were excluded in this study, which could influence the ability to discover the relationship between them. However, this study is not designed to observe the relationship between medication adherence and BP. Second, we used 3 months as our intervention period, which is inconsistent with our protocol. Taking time, cost, and seasonal factors into account, we decided to intervene in 3 months rather than 6 months, as patients' BP is lower in summer compared with

winter, because of which this study's results might be biased. Moreover, in a meta-analysis, the median intervention duration is 3 months, and response fatigue may be the feature of long-duration interventions [43]. In addition, the results in this study might be interpreted with caution because 69.0% of the participants were female, which might create a bias for this study's results. Finally, we used one-way communication in this study, and future studies on two-way SMS text messaging are suggested. According to a meta-analysis of 8 randomized trials [47], two-way SMS text messaging is related to substantially improved medication adherence compared with one-way SMS text messaging. Interactive SMS text messaging requires further inquiry and provides an opportunity for patients to get support from health care providers about healthy lifestyles, which can be translated into a long-term behavior change.

Despite these caveats, to our knowledge, this was the first study to evaluate the combined effect of SMS text messaging and consultation on patients with hypertension in the People's Republic of China. Given the rapidly changing lifestyle, characterized by high-salt and high-fat diets as well as low levels of physical activity, the prevalence of hypertension will continue to increase in China [48], and this study provides evidence to help manage the situation. The strength of this study is the combined intervention, randomized design, and use of SMS text messaging, which is convenient, inexpensive, and has a high penetration rate in society. Moreover, this study was conducted in CHCs, which represent the real-world setting; therefore, the results of this study can be used on a large scale.

Conclusions

We found significant improvements in medication adherence, SBP, and knowledge about hypertension but no difference in DBP between the intervention and control groups. The results of this study provided valuable insight and evidence on the design of intervention studies in developing countries and demonstrated that it is possible to improve BP control in community settings by using our intervention model. It is important to maximize the potential of SMS text messaging and consultation by increasing the feasibility and acceptance of mobile interventions, and future studies aimed at a cost-effectiveness analysis of this intervention model can also be conducted.

Acknowledgments

This study was funded by the Natural Science Foundation of China. The authors in this study are responsible for study design, data collection, data analyses, and writing the paper. The authors are grateful to the patients, nurses, physicians for taking part in this study. The MMAS (8-item) content, name, and trademarks are protected by US copyright and trademark laws. Permission for use of the scale and its coding is required. A license agreement is available from Donald E Morisky, ScD, ScM, MSPH, 14725 NE 20th St Bellevue, WA 98007, USA; dmorisky@gmail.com.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 492 KB - jmir_v22i5e16019_app1.pdf]

References

1. Špinar J. Hypertension and ischemic heart disease. Cor Vasa 2012;54(11):e433-e438 [FREE Full text] [doi: 10.1016/j.crvasa.2012.11.002]
2. Egan BM, Kjeldsen SE, Grassi G, Esler M, Mancia G. The global burden of hypertension exceeds 1.4 billion people: should a systolic blood pressure target below 130 become the universal standard? J Hypertens 2019 Jun;37(6):1148-1153. [doi: 10.1097/HJH.0000000000002021] [Medline: 30624370]
3. Benjamin EJ, Muntner P, Alonso A, Bittencourt MS, Callaway CW, Carson AP, American Heart Association Council on EpidemiologyPrevention Statistics CommitteeStroke Statistics Subcommittee. Heart disease and stroke statistics-2019 update: a report from the American Heart Association. Circulation 2019 Mar 05;139(10):e56-e528. [doi: 10.1161/CIR.0000000000000659] [Medline: 30700139]
4. World Health Organization. A global brief on Hypertension URL: http://ish-world.com/downloads/pdf/global_brief_hypertension.pdf [accessed 2019-10-21]
5. Mohsen Ibrahim M. Hypertension in developing countries: a major challenge for the future. Curr Hypertens Rep 2018 May 01;20(5):38. [doi: 10.1007/s11906-018-0839-1] [Medline: 29717393]
6. Wang Z, Chen Z, Zhang L, Wang X, Hao G, Zhang Z, China Hypertension Survey Investigators. Status of hypertension in China: results from the China hypertension survey, 2012-2015. Circulation 2018 May 29;137(22):2344-2356. [doi: 10.1161/CIRCULATIONAHA.117.032380] [Medline: 29449338]
7. Li L, Rao K, Kong L, Yao C, Xiang H, Zhai F, Technical Working Group of China National NutritionHealth Survey. [A description on the Chinese national nutrition and health survey in 2002]. Zhonghua Liu Xing Bing Xue Za Zhi 2005 Jul;26(7):478-484. [Medline: 16334996]

8. Lewington S, Clarke R, Qizilbash N, Peto R, Collins R, Prospective Studies Collaboration. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. Lancet 2002 Dec 14;360(9349):1903-1913. [doi: 10.1016/s0140-6736(02)11911-8] [Medline: 12493255]
9. Lip YH, Hall JE. Comprehensive Hypertension. Philadelphia: Mosby; 2007.
10. Bobrow K, Brennan T, Springer D, Levitt NS, Rayner B, Namane M, et al. Efficacy of a text messaging (SMS) based intervention for adults with hypertension: protocol for the StAR (SMS Text-message Adherence suppoRt trial) randomised controlled trial. BMC Public Health 2014 Jan 11;14:28 [FREE Full text] [doi: 10.1186/1471-2458-14-28] [Medline: 24410738]
11. Fernandez-Lazaro CI, Adams DP, Fernandez-Lazaro D, Garcia-González JM, Caballero-Garcia A, Miron-Canelo JA. Medication adherence and barriers among low-income, uninsured patients with multiple chronic conditions. Res Social Adm Pharm 2019 Jun;15(6):744-753. [doi: 10.1016/j.sapharm.2018.09.006] [Medline: 30241872]
12. Butler RJ, Davis TK, Johnson WG, Gardner HH. Effects of nonadherence with prescription drugs among older adults. Am J Manag Care 2011 Feb;17(2):153-160 [FREE Full text] [Medline: 21473664]
13. MacLaughlin EJ, Raehl CL, Treadway AK, Sterling TL, Zoller DP, Bond CA. Assessing medication adherence in the elderly: which tools to use in clinical practice? Drugs Aging 2005;22(3):231-255. [doi: 10.2165/00002512-200522030-00005] [Medline: 15813656]
14. Foreman KF, Stockl KM, Le LB, Fisk E, Shah SM, Lew HC, et al. Impact of a text messaging pilot program on patient medication adherence. Clin Ther 2012 May;34(5):1084-1091. [doi: 10.1016/j.clinthera.2012.04.007] [Medline: 22554973]
15. Abegaz TM, Shehab A, Gebreyohannes EA, Bhagavathula AS, Elnour AA. Nonadherence to antihypertensive drugs: a systematic review and meta-analysis. Medicine (Baltimore) 2017 Jan;96(4):e5641 [FREE Full text] [doi: 10.1097/MD.0000000000005641] [Medline: 28121920]
16. Ashoorkhani M, Majdzadeh R, Gholami J, Eftekhar H, Bozorgi A. Understanding non-adherence to treatment in hypertension: a qualitative study. Int J Community Based Nurs Midwifery 2018 Oct;6(4):314-323 [FREE Full text] [Medline: 30465004]
17. Vrijens B, Antoniou S, Burnier M, de la Sierra A, Volpe M. Current situation of medication adherence in hypertension. Front Pharmacol 2017;8:100 [FREE Full text] [doi: 10.3389/fphar.2017.00100] [Medline: 28298894]
18. Nieuwlaat R, Wilczynski N, Navarro T, Hobson N, Jeffery R, Keepanasseril A, et al. Interventions for enhancing medication adherence. Cochrane Database Syst Rev 2014 Nov 20(11):CD000011. [doi: 10.1002/14651858.CD000011.pub4] [Medline: 25412402]
19. Leon N, Surender R, Bobrow K, Muller J, Farmer A. Improving treatment adherence for blood pressure lowering via mobile phone SMS-messages in South Africa: a qualitative evaluation of the SMS-text Adherence SuppoRt (StAR) trial. BMC Fam Pract 2015 Jul 3;16:80 [FREE Full text] [doi: 10.1186/s12875-015-0289-7] [Medline: 26137844]
20. Bobrow K, Farmer AJ, Springer D, Shanyinde M, Yu L, Brennan T, et al. Mobile phone text messages to support treatment adherence in adults with high blood pressure (SMS-Text Adherence Support [StAR]): a single-blind, randomized trial. Circulation 2016 Feb 09;133(6):592-600 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.017530] [Medline: 26769742]
21. Farmer A, Bobrow K, Leon N, Williams N, Phiri E, Namadingo H, et al. Mobile messaging support versus usual care for people with type 2 diabetes on glycemic control: protocol for a multicenter randomized controlled trial. JMIR Res Protoc 2019 May 30;8(6):e12377 [FREE Full text] [doi: 10.2196/12377] [Medline: 31199346]
22. Adler AJ, Martin N, Mariani J, Tajer CD, Owolabi OO, Free C, et al. Mobile phone text messaging to improve medication adherence in secondary prevention of cardiovascular disease. Cochrane Database Syst Rev 2017 Apr 29;4:CD011851 [FREE Full text] [doi: 10.1002/14651858.CD011851.pub2] [Medline: 28455948]
23. Ruan Y, Xiao X, Chen J, Li X, Williams AB, Wang H. Acceptability and efficacy of interactive short message service intervention in improving HIV medication adherence in Chinese antiretroviral treatment-naïve individuals. Patient Prefer Adherence 2017;11:221-228 [FREE Full text] [doi: 10.2147/PPA.S120003] [Medline: 28228652]
24. Khorassani F, Tellier S, Tsapepas D. Pharmacist's role in improving medication adherence in transplant recipients with comorbid psychiatric disorders. J Pharm Pract 2019 Oct;32(5):568-578. [doi: 10.1177/0897190018764074] [Medline: 29554846]
25. Azhar S, Hassali MA, Ibrahim MI, Ahmad M, Masood I, Shafie AA. The role of pharmacists in developing countries: the current scenario in Pakistan. Hum Resour Health 2009 Jul 13;7:54 [FREE Full text] [doi: 10.1186/1478-4491-7-54] [Medline: 19594916]
26. Walia K, Ohri VC, Mathai D, Antimicrobial Stewardship Programme of ICMR. Antimicrobial stewardship programme (AMSP) practices in India. Indian J Med Res 2015 Aug;142(2):130-138 [FREE Full text] [doi: 10.4103/0971-5916.164228] [Medline: 26354210]
27. Pastakia SD, Vincent WR, Manji I, Kamau E, Schellhase EM. Clinical pharmacy consultations provided by American and Kenyan pharmacy students during an acute care advanced pharmacy practice experience. Am J Pharm Educ 2011 Apr 11;75(3):42 [FREE Full text] [doi: 10.5688/ajpe75342] [Medline: 21655396]
28. Szkiladz A, Carey K, Ackerbauer K, Heelon M, Friderici J, Kopcza K. Impact of pharmacy student and resident-led discharge counseling on heart failure patients. J Pharm Pract 2013 Dec;26(6):574-579. [doi: 10.1177/0897190013491768] [Medline: 23797199]

29. Fang R, Li X. Electronic messaging support service programs improve adherence to lipid-lowering therapy among outpatients with coronary artery disease: an exploratory randomised control study. J Clin Nurs 2016 Mar;25(5-6):664-671. [doi: 10.1111/jocn.12988] [Medline: 26522838]
30. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Feb 24;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
31. Guo Y, Xu Z, Qiao J, Hong YA, Zhang H, Zeng C, et al. Development and feasibility testing of an mHealth (text message and WeChat) intervention to improve the medication adherence and quality of life of people living with HIV in China: pilot randomized controlled trial. JMIR Mhealth Uhealth 2018 Sep 04;6(9):e10274 [FREE Full text] [doi: 10.2196/10274] [Medline: 30181109]
32. 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]
33. Abughosh S, Wang X, Serna O, Esse T, Mann A, Masilamani S, et al. A motivational interviewing intervention by pharmacy students to improve medication adherence. J Manag Care Spec Pharm 2017 May;23(5):549-560 [FREE Full text] [doi: 10.18553/jmcp.2017.23.5.549] [Medline: 28448784]
34. Zhai P, Li Q, Gillani AH, Hayat K, Shi L, Wang S, et al. The impact of short message services and personal consultation by pharmacy students on medication adherence and blood pressure control: study protocol for a cluster randomized trial. Patient Prefer Adherence 2019;13:627-636 [FREE Full text] [doi: 10.2147/PPA.S197858] [Medline: 31118586]
35. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
36. Morisky DE, Ang A, Krousel-Wood M, Ward HJ. Predictive validity of a medication adherence measure in an outpatient setting. J Clin Hypertens (Greenwich) 2008 May;10(5):348-354. [doi: 10.1111/j.1751-7176.2008.07572.x] [Medline: 18453793]
37. Haramiova Z, Stasko M, Hulin M, Tesar T, Kuzelova M, Morisky DM. The effectiveness of daily SMS reminders in pharmaceutical care of older adults on improving patients' adherence to antihypertensive medication (SPPA): study protocol for a randomized controlled trial. Trials 2017 Jul 18;18(1):334 [FREE Full text] [doi: 10.1186/s13063-017-2063-8] [Medline: 28720121]
38. World Health Organization. Joint FIP/WHO guidelines on good pharmacy practice: standards for quality of pharmacy services URL: https://www.who.int/medicines/areas/quality_safety/quality_assurance/FIPWHOGuidelinesGoodPharmacyPracticeTRS961Annex8.pdf [accessed 2019-10-21]
39. China Battery Expo. National guidelines for the prevention and treatment of hypertension in primary settings URL: http://www.chinabatteryexpo.com/upload/src/100113/file/2018/0226/153155373.pdf [accessed 2019-10-21]
40. Liu Z, Chen S, Zhang G, Lin A. Mobile phone-based lifestyle intervention for reducing overall cardiovascular disease risk in Guangzhou, China: a pilot study. Int J Environ Res Public Health 2015 Dec 17;12(12):15993-16004 [FREE Full text] [doi: 10.3390/ijerph121215037] [Medline: 26694436]
41. Strandberg TE, Pitkala K. What is the most important component of blood pressure: systolic, diastolic or pulse pressure? Curr Opin Nephrol Hypertens 2003 May;12(3):293-297. [doi: 10.1097/01.mnh.0000069868.94246.ef] [Medline: 12698068]
42. Jackson SL, DesRoches CM, Frosch DL, Peacock S, Oster NV, Elmore JG. Will use of patient portals help to educate and communicate with patients with diabetes? Patient Educ Couns 2018 May;101(5):956-959. [doi: 10.1016/j.pec.2017.11.004] [Medline: 29153758]
43. Thakkar J, Kurup R, Laba T, Santo K, Thiagalingam A, Rodgers A, et al. Mobile telephone text messaging for medication adherence in chronic disease: a meta-analysis. JAMA Intern Med 2016 Mar;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]
44. Varleta P, Acevedo M, Akel C, Salinas C, Navarrete C, García A, et al. Mobile phone text messaging improves antihypertensive drug adherence in the community. J Clin Hypertens (Greenwich) 2017 Dec;19(12):1276-1284 [FREE Full text] [doi: 10.1111/jch.13098] [Medline: 28941056]
45. Muntner P, Levitan EB, Joyce C, Holt E, Mann D, Oparil S, et al. Association between antihypertensive medication adherence and visit-to-visit variability of blood pressure. J Clin Hypertens (Greenwich) 2013 Feb;15(2):112-117 [FREE Full text] [doi: 10.1111/jch.12037] [Medline: 23339729]
46. Monnette A, Zhang Y, Shao H, Shi L. Concordance of adherence measurement using self-reported adherence questionnaires and medication monitoring devices: an updated review. Pharmacoeconomics 2018 Jan;36(1):17-27. [doi: 10.1007/s40273-017-0570-9] [Medline: 28895104]
47. Wald DS, Butt S, Bestwick JP. One-way versus two-way text messaging on improving medication adherence: meta-analysis of randomized trials. Am J Med 2015 Oct;128(10):1139.e1-1139.e5. [doi: 10.1016/j.amjmed.2015.05.035] [Medline: 26087045]

48. Guo QH, Zhang YQ, Wang JG. Asian management of hypertension: current status, home blood pressure, and specific concerns in China. J Clin Hypertens (Greenwich) 2019 Oct 17:- (forthcoming) [FREE Full text] [doi: 10.1111/jch.13687] [Medline: 31622005]

Abbreviations

BP: blood pressure
CHC: community health care center
DBP: diastolic blood pressure
GPP: Guidelines for Good Pharmacy Practice
MMAS-8: 8-item Morisky Medication Adherence Scale
SBP: systolic blood pressure

Edited by G Eysenbach; submitted 27.08.19; peer-reviewed by X Wang, B Smith; comments to author 15.10.19; revised version received 11.11.19; accepted 26.01.20; published 20.05.20.

Please cite as:
Zhai P, Hayat K, Ji W, Li Q, Shi L, Atif N, Xu S, Li P, Du Q, Fang Y
Efficacy of Text Messaging and Personal Consultation by Pharmacy Students Among Adults With Hypertension: Randomized Controlled Trial
J Med Internet Res 2020;22(5):e16019
URL: https://www.jmir.org/2020/5/e16019
doi:10.2196/16019
PMID:32432556

Original Paper

A Personalized eHealth Intervention for Lifestyle Changes in Patients With Cardiovascular Disease: Randomized Controlled Trial

Eva Rosalinde Broers[1,2*], PhD; Willem Johan Kop[2*], Prof Dr; Johan Denollet[2*], Prof Dr; Jos Widdershoven[1,2*], Prof Dr, MD; Mart Wetzels[3*], PhD; Idowu Ayoola[3*], PhD; Jordi Piera-Jimenez[4*], MSc; Mirela Habibovic[1,2*], PhD

[1]Department of Cardiology, Elisabeth-Tweesteden Hospital, Tilburg, Netherlands

[2]Department of Medical and Clinical Psychology, Tilburg University, Tilburg, Netherlands

[3]University of Technology Eindhoven, Eindhoven, Netherlands

[4]Badalona Serveis Assistencials, Badalona, Spain

[*]all authors contributed equally

Corresponding Author:
Mirela Habibovic, PhD
Department of Medical and Clinical Psychology
Tilburg University
Tilburg
Netherlands
Phone: 31 13 466 4020
Email: m.habibovic@tilburguniversity.edu

Abstract

Background: Behavior change methods involving new ambulatory technologies may improve lifestyle and cardiovascular disease outcomes.

Objective: This study aimed to provide proof-of-concept analyses of an intervention aiming to increase (1) behavioral flexibility, (2) lifestyle change, and (3) quality of life. The feasibility and patient acceptance of the intervention were also evaluated.

Methods: Patients with cardiovascular disease (N=149; mean age 63.57, SD 8.30 years; 50/149, 33.5% women) were recruited in the *Do Cardiac Health Advanced New Generation Ecosystem (Do CHANGE)* trial and randomized to the *Do CHANGE* intervention or *care as usual* (CAU). The intervention involved a 3-month behavioral program in combination with ecological momentary assessment and intervention technologies.

Results: The intervention was perceived to be feasible and useful. A significant increase in lifestyle scores over time was found for both groups ($F_{2,146.6}$=9.99; P<.001), which was similar for CAU and the intervention group ($F_{1,149.9}$=0.09; P=.77). Quality of life improved more in the intervention group (mean 1.11, SD 0.11) than CAU (mean −1.47, SD 0.11) immediately following the intervention (3 months), but this benefit was not sustained at the 6-month follow-up (interaction: P=.02). No significant treatment effects were observed for behavioral flexibility ($F_{1,149.0}$=0.48; P=.07).

Conclusions: The Do CHANGE 1 intervention was perceived as useful and easy to use. However, no long-term treatment effects were found on the outcome measures. More research is warranted to examine which components of behavioral interventions are effective in producing long-term behavior change.

Trial Registration: ClinicalTrials.gov NCT02946281; https://www.clinicaltrials.gov/ct2/show/NCT02946281

(J Med Internet Res 2020;22(5):e14570) doi:10.2196/14570

KEYWORDS

cardiovascular diseases; lifestyle; habits; eHealth; mHealth

Introduction

The elimination of modifiable behavioral risk factors for cardiovascular disease (eg, smoking and physical inactivity) in the general population could prevent 80% of adverse clinical outcomes [1]. In patients diagnosed with cardiovascular disease, a modest reduction in risk behaviors can decrease the mortality rates by approximately 50% [2]. However, recommended targets (eg, lifestyle and medication adherence) for secondary prevention are rarely reached [3]. To achieve sustained health behavior change, active interventions that go beyond patient education are needed [4].

Sustainable behavior changes can be enhanced by implementing a personalized patient-tailored approach [5,6]. New ambulatory technologies can now be used to provide personalized support in a low threshold, nonobtrusive, and ecologically valid manner. These devices can be used to provide feedback about ambulatory health behaviors (eg, physical activity levels), but they are not sufficient to produce long-term behavior change [7,8]. In the setting of cardiac rehabilitation, telemonitoring guidance for patients' physical activity levels was found to be feasible in the FIT@Home study, and this intervention resulted in higher patient satisfaction and trends toward lower health care costs, but not in better improvements in fitness or physical activity levels relative to standard center-based rehabilitation [8]. The impact of this intervention could potentially have been further improved if a more patient-tailored approach were added. Another study found initial support that an app using persuasive design techniques can improve biological and psychological factors in patients after cardiac rehabilitation [9]. It is therefore plausible that ambulatory assessments are likely to have better therapeutic effects when combined with prompts that promote health-related behaviors (ie, ecological momentary interventions). These new methodologies also require a deeper knowledge about patients' needs and preferences [10].

This trial (*Do Cardiac Health Advanced New Generation Ecosystem*, Do CHANGE) was specifically designed to examine this multidisciplinary approach to behavior change [11]. What is unique to this trial is that patients received the behavior change program, *Do Something Different* (DSD), which has been previously developed to change unhealthy habits through the increase of behavioral flexibility [11]. Behavioral flexibility is associated with a broad range of the behavioral repertoire, making people more open to experience and the adoption of new behaviors [12]. DSD has been evaluated in other patient samples and has shown promising results by producing health behavior change [13]. For this study, the program was adapted to meet the needs of patients with cardiovascular disease (coronary artery disease, CAD; heart failure, HF; and hypertension, HT). Hence, the aim of this study was to provide proof of concept for the behavioral intervention aiming to address (1) behavioral flexibility, (2) lifestyle change, and (3) quality of life. The feasibility and patient acceptance of the intervention were also evaluated.

Methods

Design

The Do CHANGE trial is an international (the Netherlands and Spain), multicenter, randomized controlled trial, designed to enhance lifestyle changes in patients with cardiac disease (NCT02946281). The trial findings described in this paper are the first (proof of concept and feasibility) phase of the Do CHANGE project (phase 1) and will serve as input for further development of a second phase of this randomized controlled trial (Do CHANGE, phase 2; NCT03178305). A detailed description of both phases of the Do CHANGE trial has been published previously [11]. As this trial was developed to provide information about proof of concept and feasibility, an a priori sample size calculation was not performed. For this phase, we aimed to include 150 patients across 2 countries, which is considered sufficient to give information about proof of concept and feasibility of the intervention.

Study Sample

Patients diagnosed with CAD (having experienced a myocardial infarction, percutaneous coronary intervention, angina pectoris, or coronary artery bypass graft surgery), symptomatic HF (New York Heart Association class I-IV), and HT were included in the study. HT was defined as systolic blood pressure ≥140 mm Hg and/or diastolic blood pressure ≥90 mm Hg on two different measurements spaced 1 to 2 min apart and after 3 to 5 min in a sitting position. The values of the second measure were used. HF patients were included if they had a diagnosis of systolic or diastolic HF and the presence of HF symptoms.

Patients were recruited at Badalona Serveis Assistencials (Badalona, Spain) and Elisabeth-TweeSteden Hospital (Tilburg, the Netherlands). The study was approved by the medical ethics committees of the participating hospitals and was conducted in accordance with the Helsinki Declaration.

The inclusion criteria were as follows: (1) a primary diagnosis of CAD, HF, or HT; (2) aged 18 to 75 years; (3) having ≥2 of the following risk factors: positive family history, increased cholesterol, smoking, diabetes, sedentary lifestyle, and/or psychosocial risk factors; (4) sufficient knowledge of the country's native language; (5) access to the internet at home; and (6) having a smartphone compatible with the apps used in the study.

The exclusion criteria were as follows: (1) life expectancy <1 year, (2) life-threatening comorbidities (eg, malignancy), (3) history of psychiatric illness other than anxiety and/or depression, (4) significant cognitive impairments (eg, dementia), and (5) on the waiting list for heart transplantation.

Procedure

Patients meeting the inclusion criteria were approached for participation by a cardiologist or cardiac nurse. If interested, patients received information about the study in writing and orally. After 10 days, patients were contacted to inquire about their participation. If the patient indicated that they wanted to participate, a face-to-face appointment was scheduled at the hospital. Patients were asked to sign an informed consent

document and were provided with the first set of questionnaires (baseline). After filling in the questionnaires, patients were randomized. Patients in the intervention group received information about the intervention and the use of associated devices (see the *Intervention* section). The following day, patients in the intervention group were contacted by telephone to check that the devices were installed correctly and that the system was functional.

Patients received the intervention for 3 months. Follow-up questionnaires were sent at 3 and 6 months. Patients returned the devices after completion of the intervention (ie, after 3 months).

Randomization and Blinding

Patients were randomized (1:1) after completing the baseline questionnaires. Randomization sequences were computer generated and individually sealed before recruitment started. After completing the questionnaires, one sealed envelope was drawn by the research assistant containing the group allocation. Owing to the nature of the behavioral monitoring aspects of the study, blinding health care providers or participants to the treatment condition was not possible, whereas the initial analyses of the study outcomes were analyzed without knowledge of the treatment allocation.

Intervention

Do Cardiac Health Advanced New Generated Ecosystem Intervention Versus Care as Usual

Behavior Change Technique

Patients randomized to the intervention group received a 3-month behavior change program, *DSD*, which was provided via text messages on patients' mobile phones. The DSD program that was used aims to change unhealthy habits through the increase of behavioral flexibility [12]. This is achieved by disrupting patients' daily behavioral routine for a short period (few seconds) with behavioral prompts, which are referred to as *Do's* (eg, "EXPLORE MORE DAY. Today instead of going the same old way, take a different route. Look around, spot 10 things you wouldn't see on your usual journey") and are provided through patients' mobile phones. These messages challenge patients to do something different and get out of their comfort zone. They have been developed by a multidisciplinary team, including cardiologists and psychologists, to make sure that the Do's apply to the patient population and are thus related to their daily behaviors/needs. Patients received a total of 32 Do's during the 3-month intervention period (2-3 Do's every week). DSD has been evaluated in other patient samples and has shown promising results with respect to behavior change [13]. For this trial, the program was adapted to the cardiac population with slight differences in the program depending on patients' primary diagnosis (eg, CAD, HF, HT), as the preferred health behaviors may vary depending on the diagnosis. For example, because of disease-specific symptoms, advice regarding fluid intake was taken into account within the program only for patients with HF. More details regarding the DSD program are provided in the previously published design paper of this project [11].

Technological Tools

In addition to the DSD program, to obtain objective measures on patients' physical functioning, all patients received a blood pressure monitor, the Moves app (ProtoGeo, Helsinki; to register the GPS location), and the CarePortal (Docobo Ltd, Leatherhead; eg, a home monitoring device measuring daily symptoms and electrocardiogram). Owing to the disease-specific reasons, patients diagnosed with HF also received a weight scale, as daily weight monitoring is of importance in this subgroup. Data obtained from these devices will not be included in this analysis. This manuscript will focus on the primary outcome measures related to lifestyle parameters and patient-reported outcomes, which were derived from validated questionnaires (see the study by Habibović et al [11] for a description of primary and secondary outcomes).

Control Group

Patients randomized to the care as usual (CAU) group received the treatment as usual and were only provided with the validated questionnaires at baseline and at 3 and 6 months. These patients did not receive devices for ambulatory monitoring measures.

Measures

Questionnaires

Primary outcomes

Behavioral Flexibility

Behavioral flexibility was measured using the *DSD* questionnaire from scale items designed for this study [11]. This scale contains 30 different descriptions of behavior coupled in 15 pairs of opposites (Multimedia Appendix 1). Patients were asked at each measurement point to select the behaviors that best describe them (eg, *gentle* or *firm*). On the basis of a formula, the behavioral flexibility for each participant at each time point was calculated as outlined below.

Every addition of behavior raises the score as well as when both of a pair of opposite behaviors are added. For example, definite, systematic, trusting, predictable, and unpredictable are selected. All these selected behaviors raise the flexibility score. However, because predictable and unpredictable are each other's opposites, these are added to the formula again and increase the flexibility score even more. The model interprets this seemingly contradictory behavior as evidence of flexibility: based on what a given situation demands, the person can use different reactions and thus be more flexible. The total score can range from 0 to 100. The internal consistency in this sample was considered acceptable (Cronbach alpha=.67 to .76).

Lifestyle

The Health-Promoting Lifestyle Profile questionnaire was administered to assess health-promoting lifestyle habits [14]. This survey evaluates whether the subjective perception of patients regarding their lifestyle is changed and consists of 52 items (eg, "Eat 6-11 servings of bread, cereal, rice, and pasta each day") in total. Each item can be answered on a 4-point Likert scale, ranging from 1 (never) to 4 (routinely). The total

score can therefore range from 52 to 208, with a higher score indicating a better lifestyle. Furthermore, the questionnaire includes 6 different subscales that each cover a health promotion lifestyle domain (ie, Physical Activity, Spiritual Growth, Health Responsibility, Interpersonal Relationships, Nutrition, and Stress Management). The internal consistency was considered as excellent in this sample (Cronbach alpha=.88 to .90).

Quality of Life

To administer changes in the quality of life, participants completed the World Health Organization Quality of Life—BREF (WHOQOL-BREF) [15]. The WHOQOL-BREF is considered a reliable, generic multidimensional quality of life measure and consists of 26 items in total. Two items refer to the facet's overall quality of life and general health, whereas the abiding 24 items reflect 4 different domains (ie, physical health, psychological health, social relationships, and environment). The internal consistency in this sample was excellent (Cronbach alpha=.89 to .90).

Perceived Usefulness and Acceptance

The Unified Theory of Acceptance 2 (UTAUT2) scale [16] was administered to assess the perceived usefulness and acceptance of the tools that were used in the intervention. Mean scores on 8 subscales are provided, namely, (1) Performance Expectancy, (2) Effort Expectancy, (3) Social Influence, (4) Facilitating Conditions, (5) Hedonic Motivation, (6) Habit, and (7) Behavioral Intention. The initial subscale, *Price Value*, of the UTAUT2 was not included, as the cost per individual for the intervention could not be estimated. The total score per subscale can range from 4 to 20, with a higher score indicating higher usefulness and acceptance [16]. The internal consistency in this sample was excellent (Cronbach alpha=.89).

Client Satisfaction Questionnaire

To assess the satisfaction of the patients about the ecosystem, the 8-item Client Satisfaction Questionnaire [17] was used. This self-administered questionnaire is a general scale that consists of 8 Likert scale items (eg, "To what extent has our program met your needs?") ranging from 0 to 4, with response descriptors that vary. The overall score can range from 8 to 32, with a higher score indicating a higher satisfaction. The internal consistency was rated as excellent (Cronbach alpha=.92) [17].

Other Questionnaires Included in the Model

Type D Scale (Distress Scale-14)

Type D personality was assessed using the Type D scale (Distress Scale-14) [18]. This 14-item questionnaire consists of 2 subscales with seven 5-point Likert scale items each, ranging from 0 (false) to 4 (true). Total scores on both subscales range from 0 to 28. The 2 subscales represent the characteristics of negative affectivity (NA; eg, the tendency to experience negative emotions across time and situations) and social inhibition (SI; eg, the tendency not to express feelings). When scoring ≥10 on both subscales, patients were classified as Type D. With a reported Cronbach alpha value of .86 and .88, respectively, the internal consistency of SI and NA are considered as satisfactory [18].

The 7-Item Generalized Anxiety Disorder Scale

To gauge self-administered symptoms of anxiety, the 7-item Generalized Anxiety Disorder scale was administered [19]. The questionnaire is comprised of 7 items (eg, "Feeling afraid as if something terrible might happen") that can be answered on a 4-point Likert scale, ranging from 0 (not at all) to 3 (almost every day). To get an indication of anxiety symptom severity, the total score (range from 0 to 21) can be used. A higher score implies higher levels of anxiety. The internal consistency was considered excellent (Cronbach alpha=.92) [19].

Nine-Item Patient Health Questionnaire

Depressive symptoms were administered at each time point by using the 9-item Patient Health Questionnaire (PHQ-9) [20]. This self-report questionnaire consists of 9 items in total (eg, *feeling down*, *depressed*, or *hopeless*), each evaluated on a 4-point Likert scale (ie, not at all, several days, and nearly every day). The total score ranges from 0 to 27, with a higher score as an indication of worse depression symptom severity. The internal consistency was considered excellent (Cronbach alpha=.90) [21].

Demographic and Clinical Data

Demographic characteristics (eg, age, sex, marital status, working status, level of education, and smoking behavior) were obtained by patients' self-report. Clinical data (eg, comorbidities; prescribed cardiac medication; prescribed psychotropic medication; left ventricular ejection fraction; history of coronary artery bypass grafting; history of percutaneous coronary intervention; and resting heart rate and systolic and diastolic blood pressure measured at the most recent outpatient visit) were obtained from the medical record.

Statistical Analyses

Categorical variables were compared using chi-square tests, and continuous variables were compared using a two-tailed *t* test for independent samples. To evaluate the treatment effectiveness, based on intention-to-treat, a univariate and multivariate Linear Mixed Model analysis was performed. Multivariable analyses were adjusted for age, sex, education, site of inclusion (Badalona Serveis Assistencials [BSA] or Elisabeth-TweeSteden Hospital [ETZ]), primary diagnosis, Charlson comorbidity index scores, Type D personality, baseline anxiety scores, and baseline depression scores. *F* values with two-sided *P* values were reported for main and interaction effects (group×time). For the estimated fixed effects, beta coefficients with two-sided *P* values were reported. Data were analyzed using the SPSS software package (version 24).

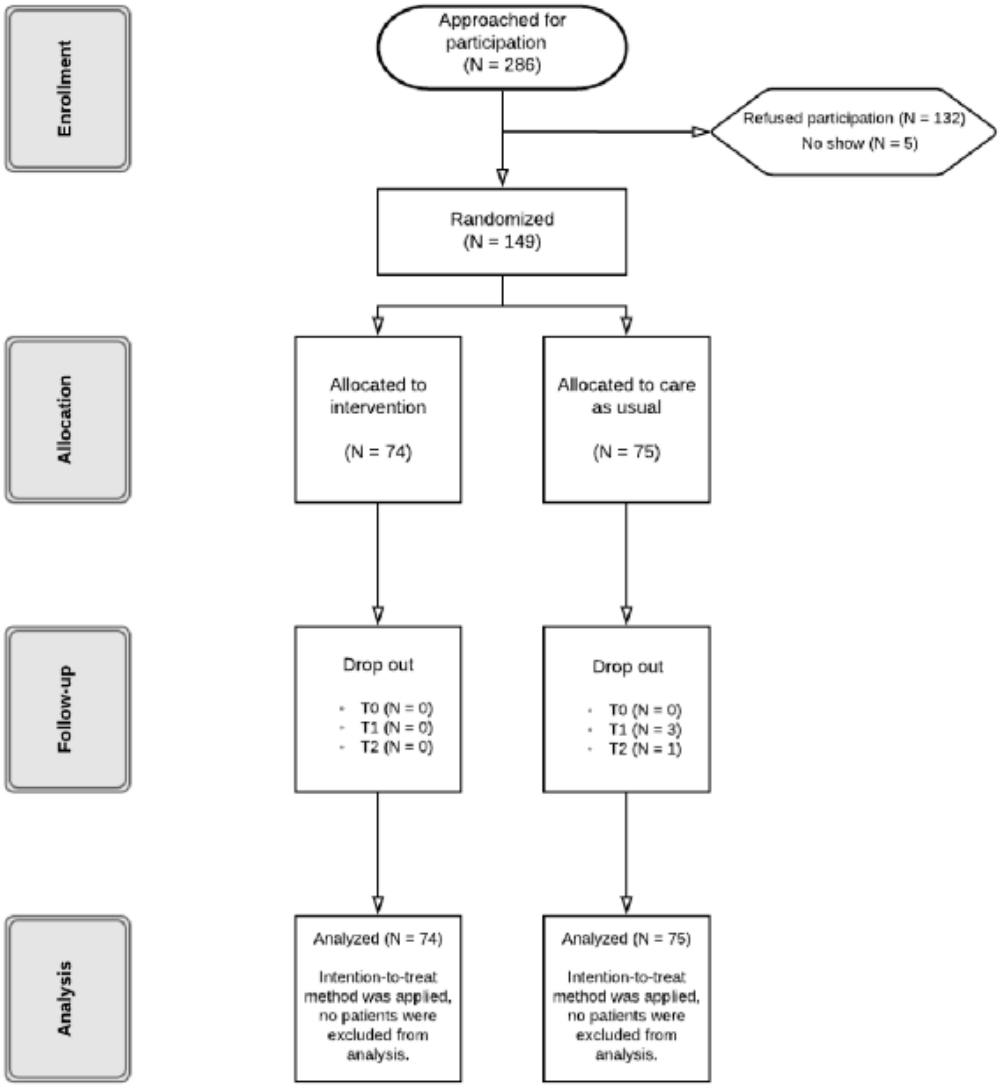
Results

Sample

The data collection took place between January 2017 and September 2017. In total, 286 eligible patients were approached for participation, of which 132/286 (46.1%) patients refused to participate. An additional 5/286 (1.7%) participants did not show up or declined participation. The reasons for refusal included that it would be too time consuming, they did not want to be confronted about their heart disease every day, and they

were reluctant to use technology. A total of 149 (response rate: 149/286; 52.0%) patients were enrolled. Enrollment per study site was as follows: BSA randomized a total of 74 (intervention: n=37 and CAU: n=37) patients; ETZ randomized 75 patients (intervention: n = 37 and CAU: n=38). Of the total sample, 4 participants within the CAU condition dropped out, as they did not receive the intervention and were therefore not willing to continue. Of the patients randomized to the intervention condition, 82% (61/74) reported having completed the entire 3-month program. Overall, 97.3% (145/149) of the participants completed the follow-up assessments. Figure 1 presents the flowchart of patient recruitment.

Figure 1. Flow diagram of patient recruitment.



Baseline Characteristics

The mean age of the total sample was 63.6 (SD 8.3) years, and 66% (99/149) were men. There were significant differences observed in mean completed education in years between the intervention group (mean 14.3, SD 6.2) and the CAU group (mean 11.8, SD 7.9; *P*=.03). This means that patients in the intervention group completed more years of education than those in the CAU group. Furthermore, a significant difference between the 2 groups was found on the mean PHQ-9 baseline scores, with a higher mean score on depressive symptoms in the CAU group (mean 3.61, SD 3.6 vs mean 5.56, SD 4.17; *P*=.003). No other differences were found between the intervention and CAU groups. Table 1 presents an overview of the baseline characteristics of this sample.

Table 1. Baseline patient characteristics of the total sample.

Variable	Total	Do Cardiac Health Advanced New Generation Ecosystem intervention (N=74)	Care as usual (N=75)	*P* value
Site of allocation, n (%)				
Badalona Serveis Assistencials	74 (49.7)	37 (50)	37 (50)	N/A[a]
Elisabeth-TweeSteden Hospital	75 (50.3)	37 (49.3)	38 (50.7)	N/A
Total	149 (100)	74 (49.7)	75 (50.3)	N/A
Demographics				
Age (years), mean (SD)	63.57 (8.30)	63.26 (8,35)	63.88 (8.30)	.65
Gender (male), n (%)	99 (66.4)	52 (70.3)	47 (62.7)	.42
Education (years), mean (SD)	13.03 (7.22)	14.30 (6.24)	11.79 (7.91)	.03
Marital status (partner), n (%)	118 (79.2)	61 (82.4)	57 (76.0)	.44
Working status (working), n (%)	55 (36.9)	28 (37.8)	27 (36.0)	.95
Smoking (yes), n (%)	27 (18.1)	10 (13.5)	17 (22.7)	.29
Clinical				
Diagnosis heart failure, n (%)	36 (24.2)	21 (28.4)	15 (20.0)	.96
Diagnosis hypertension, n (%)	73 (49.0)	38 (51.4)	35 (46.7)	>.99
Diagnosis coronary artery disease, n (%)	40 (26.8)	15 (20.3)	33 (33.3)	.07
Charlson comorbidity index, mean (SD)	1.14 (0.95)	1.01 (0.88)	1.27 (1.00)	.11
Systolic blood pressure (baseline), mean (SD)	138.02 (19.71)	135.00 (20.89)	141.00 (18.13)	.06
Diastolic blood pressure (baseline), mean (SD)	79.27 (10.01)	78.76 (10.33)	79.77 (9.87)	.54
Heart rate (rest), mean (SD)	69.41 (11.97)	69.95 (14.43)	68.88 (11.57)	.59
Psychological				
Patient Health Questionnaire-9, mean (SD)	4.59 (4.00)	3.61 (3.60)	5.56 (4.17)	.003
Generalized Anxiety Disorder-7, mean (SD)	4.03 (4.37)	3.35 (4.07)	4.69 (4.59)	.06
Type D personality (yes), n (%)	38 (25.5)	20 (27.0)	18 (24.0)	.67

[a]N/A: not applicable.

Intervention Effects

Behavioral Flexibility

The univariate analysis on behavioral flexibility scores (including group, time, and group×time) revealed no significant main effects for group ($F_{1,148.93}$=3.42; *P*=.07) or time ($F_{2,146.82}$=1.69; *P*=.18) and group×time interaction ($F_{2,146.82}$=1.09; *P*=.34). After adjusting for covariates (as previously described), main effects for time ($F_{1,146.81}$=1.74; *P*=.18), group ($F_{1,149.00}$=0.48; *P*=.07), and group×time ($F_{2,146.81}$=1.13; *P*=.33) remained nonsignificant (Figure 2). These findings indicate that behavioral flexibility scores did not significantly change over time, and that there were no differences between the 2 groups. With regard to covariates included in the model, the estimated fixed effects of HT (β=−6.07; *P*=.01) and CAD (β=−5.57; *P*=.02) were significantly associated with lower levels of behavioral flexibility scores (across all time points). In addition, the site of recruitment was associated with behavioral flexibility scores, with only patients from Spain showing an increase in behavioral flexibility over time (β=6.31; *P*<.01) when compared with those in the Netherlands (see Table 2).

Figure 2. Mean scores of intervention and care as usual group on primary outcome measures.

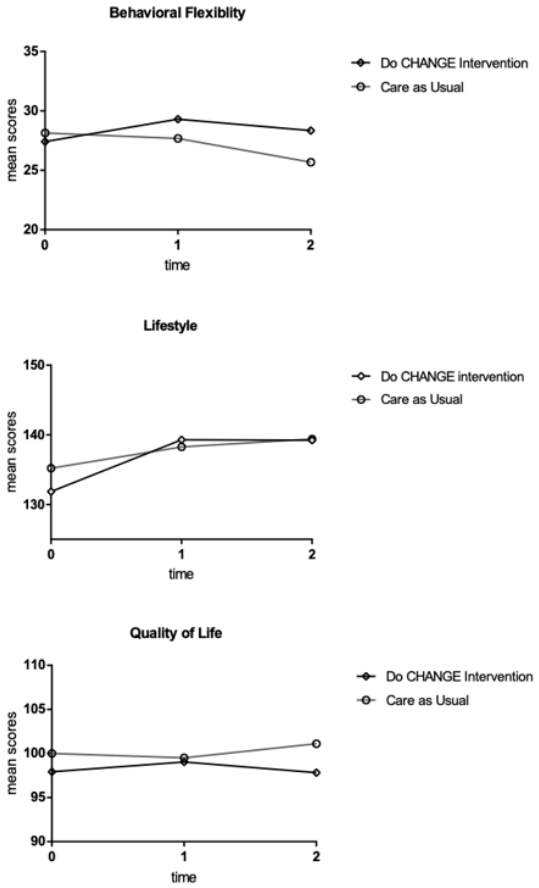


Figure 3. Standalone Equation 1.

$$\left|100\% \times \left(\frac{number\ of\ behaviors\ selected}{30} + \frac{number\ of\ opposite\ pairs\ selected}{15}\right)\right.$$

Table 2. Estimates of fixed effects from multivariable linear mixed models on the main outcome measures at baseline and at 3- and 6-month follow-up.

Multivariable linear mixed model	Behavioral flexibility			Lifestyle			Quality of life		
	Estimate	SE	*P* value	Estimate	SE	*P* value	Estimate	SE	*P* value
Model 1: unadjusted model									
Intervention group	4.70	2.07	.03	2.18	3.52	.54	0.51	1.89	.79
Model 2: covariates adjusted model									
Intervention group	2.91	1.99	.15	0.46	3.27	.89	−2.89	1.69	.09
Hypertension[a]	−6.07	2.27	.01	−3.89	3.72	.30	−3.40	2.07	.10
Coronary artery disease[a]	−5.57	2.29	.02	3.62	3.74	.33	−2.02	2.08	.33
Study site (Spain)	6.31	1.84	<.01	12.86	3.01	<.01	4.62	1.67	<.01
Sex (male)	2.36	1.79	.19	−1.66	2.93	.57	2.66	1.63	.11
Type D	−0.28	2.14	.90	8.33	13.47	.54	−10.85	7.48	.15
Age (years)	0.08	0.09	.42	−0.04	0.16	.80	0.07	0.09	.45
Higher Charlson comorbidity index	−1.73	1.00	.08	−1.09	1.63	.51	−1.76	0.90	.05
Higher education	−0.01	0.12	.95	0.08	0.19	.66	0.25	0.11	.02
Anxiety	-0.12	0.44	.79	−1.38	0.72	.06	-0.90	0.40	.03
Depression	−1.77	1.13	.12	−4.17	1.86	.03	−3.63	1.03	<.01

[a]Compared with the main diagnosis of heart failure.

Lifestyle

A univariate analysis showed no significant effect for group ($F_{1,434.91}$=0.91; *P*=.34) or group×time ($F_{2,282.73}$=0.39; *P*=.68). These findings present that, without the addition of possible confounding variables, the intervention and CAU group did not differ. In addition, no interaction effect between allocation to group and time was found. However, a significant improvement for both groups on overall reported lifestyle behavior was found ($F_{2,282.73}$=4.28; *P*=.02). When adjusting for covariates in the multivariable analysis, this improvement remained significant ($F_{2,146.63}$=9.99; *P*<.001). As shown in Figure 2, both groups reported improvements in lifestyle behavior over time. The effects of interaction ($F_{2,147.02}$=1.36; *P*=.26) and allocation to group ($F_{1,149.90}$=0.09; *P*=.77) remained nonsignificant in the adjusted models. This indicated that no effect of the intervention on healthy lifestyle behavior was found. The estimated fixed effect of depression (β=−4.17; *P*=.03) was negatively associated with lifestyle promoting behavior, indicating that patients who score higher on the depression scale report lower healthy lifestyle behaviors. Patients from Spain showed an increase in lifestyle behavior (β=12.86; *P*<.01), in comparison with those in the Netherlands (see Table 2).

Quality of Life

The results of the univariate analysis of the quality of life total scores showed an interaction effect between time and group ($F_{2,146.40}$=4.22; *P*=.02). This finding indicates that the mean scores on quality of life of the intervention and CAU groups have different slopes over time: the intervention group showed a small improvement in the quality of life after 3 months, whereas the CAU group reported a small decline in the quality of life (mean improvement 1.11, SD .11 vs mean −1.47, SD .11). Both groups stabilized to baseline level after 6 months. The interaction effect remained significant after adding the covariates in the multilevel analysis ($F_{2,146.52}$=4.29; *P*=.02; Figure 2), suggesting a significant, positive effect of the intervention on self-reported quality of life in the first 3 months. The estimated fixed effects of higher levels of education (β=.25; *P*=.02) and being recruited in Spain (β=4.62; *P*=.008) compared with those in the Netherlands were significantly associated with higher scores on quality of life. Lower scores were predicted by higher CCI scores (β=−1.76; *P*=.05) anxiety (β=-0.90; *P*=.03), and depression (β=−3.63; *P*<.01; see Table 2). Examining subscales of the WHOQOL revealed no specific subscale differences regarding response patterns to the intervention.

Acceptability and Satisfaction

Overall, patients in the intervention group indicated being satisfied with the intervention (mean 26.22, SD 4.82). The intervention was perceived to be useful (mean 13.88, SD 3.96) and easy to use (mean 17.07, SD 2.57). Patients did not feel social pressure to use the devices from the intervention (mean 9.85, SD 3.63) and reported to be quite satisfied with the possibilities to receive support (mean 15.44, SD 2.43) and had a neutral opinion regarding the pleasure in using the devices offered in the intervention (mean 10.63, SD 2.44). Furthermore, the intervention was integrated relatively well in patients' lives (mean 11.71, SD 3.05). However, patients indicated that they were neutral regarding the intention to use the ecosystem in the future (mean 8.40, SD 3.34).

Discussion

Principal Findings

This study aimed to provide proof of concept for the Do CHANGE behavioral intervention targeting behavioral flexibility, lifestyle change, and quality of life in cardiac patients. No significant differences between the groups were observed on behavioral flexibility and lifestyle. However, a small increase in quality of life at 3 months was observed in the intervention group, but at 6 months, no significant difference between the groups was observed. With respect to the usefulness and feasibility of the intervention, the findings of this study revealed that the ecosystem is experienced as useful, easy to use, and integrated well into the daily life of the patients. It made the participants more aware of the fact that they must undertake activities themselves to feel better. Patients also reported feeling more *safe* because health care professionals were watching along. Nonadherence is a common issue in Web-based interventions for promoting health-related behavior, and the average study results in only 50% of participants adhering to the intended intervention [22]. However, 82.4% of the patients participating in this Do CHANGE intervention condition completed the intervention, which may further indicate that the intervention was not perceived as demanding.

The findings of the study are not completely in line with previous studies in other patient populations [13]. An explanation for this discrepancy could be the fact that this was the first study implementing the concept of behavioral flexibility and thus the core Do's of the DSD program in the cardiac population. In addition, the Do's might not have been tailored enough to the patients' needs that the timing of the Do's might not have been optimal. For example, one would want a patient to receive a distractive Do at the time when the *unwanted* behavior occurs. In this trial, patients from 2 different cultures (Spain and the Netherlands) and diagnosed with different cardiac disorders (ie, HT, HF, or CAD) were enrolled. This reflects the heterogeneity of the sample, which may have affected the results. Another important area for future research is the exploration of the mediating factors that drive the interplay between behavioral flexibility, lifestyle factors, and quality of life in the setting to electronic health (eHealth) interventions.

Enrollment in the study may have increased the general awareness of lifestyle change in both groups. This awareness could unknowingly lead to the adaptation of a lifestyle, independent of the allocation to a group. Previous research in cardiac patients affirms that there is a relationship between general knowledge about cardiac risk factors and self-reported lifestyle changes in the short term [23]. Furthermore, although lifestyle change is crucial in the treatment of cardiovascular disease, there is a lack of emphasis on lifestyle change and self-care of the patient in the current health care systems [24]. Addressing self-care and lifestyle change in clinical practice is therefore warranted. The Do CHANGE trial provides (longitudinal and momentary) knowledge that can be used in the further development of personalized interventions that will help patients reach recommended lifestyle goals.

Behavioral flexibility is an important construct on which behavioral change can possibly be initiated. However, the results of this study may indicate the need for a better measurement tool, as the questionnaire that is used might not be sensitive enough to reveal significant alteration in patients' behavioral flexibility over time.

The findings on quality of life, on the other hand, are not entirely in line with previous studies in cardiac samples, which have shown that there is a decline in quality of life and generally a slight increase in anxiety and depression scores within the 3 months postcardiac event [25]. This could be explained by patients having to adapt to new behaviors after visiting the hospital and being reminded of the fact that they have a chronic illness. In this study, the intervention group received the behavioral program, which could have contributed to first an increase in their quality of life, with a slight decrease after 3 months, sustaining their baseline quality of life. After the behavioral intervention ends, the quality of life in the intervention group also goes down as the additional *support* is no longer provided.

The findings of this study must be interpreted in light of a few limitations. At baseline, the intervention group and CAU group showed some differences in mean years of completed education and mean scores on depressive symptoms. The CAU group scored significantly higher on both variables. Another limitation of this study was that the sample was rather small, in relatively good health, and clinically heterogeneous, which may have limited the possibility to find substantial effects. Although the intervention was positively evaluated by participants, half of the approached patients did refuse participation. Therefore, it can be concluded that eHealth interventions similar to those described in this study are appealing for certain subgroups of patients. Future research should focus on eHealth interventions within the cardiac population based on a larger sample with significant power that is assessed over a prolonged follow-up duration (eg, beyond 6 months) to draw firm conclusions on sustainable behavior change. The results of this study showed that depression was associated with negative behavioral and psychological outcomes, which is in line with previous findings [26]. Depressive symptoms are common in patients with cardiovascular disease [27-29] and are related to various behavioral risk factors (eg, sedentary lifestyle, unhealthy diet, alcohol overconsumption, and smoking) [30,31]; this may explain the relation between depression and lower lifestyle behavior scores. Hence, future research is needed to examine which psychological and clinical factors contributing to health behavior change and potentially address these factors during the intervention.

For clinical practice, it is important to acknowledge that technology and eHealth solutions might be the feasible way forward in meeting patient needs and initiated health behavior change. However, the findings of this study underline the importance of a personalized approach that includes the assessment of a patient's demographic, clinical, and psychological profile.

Conclusions

In conclusion, the Do CHANGE 1 intervention was perceived as useful and easy to use. However, no main effects were found on behavioral flexibility, lifestyle behavior, and quality of life. More research is warranted to examine which components of behavioral interventions, and in which patients, are effective in producing long-term behavior change.

Acknowledgments

The authors would like to thank all the patients for participating and the health care professionals in the participating hospitals for making this work possible. In addition, the authors would like to thank all the students/research assistants for their help with the patient recruitment and data management. Finally, the authors want to thank the consortium members of the Do CHANGE project. This study was funded by the European Commission's Horizon 2020 program (grant number: 463735). For this project, the Do CHANGE team received funding for research and innovation from the European Union. One startup (Onmi) and 2 small- and medium-sized enterprises (DSD, Docobo Ltd) are supported financially to develop their products.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Consolidated Standards of Reporting Trials of Electronic and Mobile HEalth Applications and onLine TeleHealth checklist.
[PNG File , 73 KB - jmir_v22i5e14570_app1.PNG]

Multimedia Appendix 2
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 507 KB - jmir_v22i5e14570_app2.pdf]

References

1. Piepoli M, Hoes A, Agewall S, Albus C, Brotons C, Catapano A, ESC Scientific Document Group. 2016 European Guidelines on cardiovascular disease prevention in clinical practice: The Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of 10 societies and by invited experts)Developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). Eur Heart J 2016 Aug 1;37(29):2315-2381. [doi: 10.1093/eurheartj/ehw106] [Medline: 27222591]
2. Chow CK, Jolly S, Rao-Melacini P, Fox KA, Anand SS, Yusuf S. Association of diet, exercise, and smoking modification with risk of early cardiovascular events after acute coronary syndromes. Circulation 2010 Feb 16;121(6):750-758. [doi: 10.1161/CIRCULATIONAHA.109.891523] [Medline: 20124123]
3. Kotseva K, Wood D, de Bacquer D, de Backer G, Rydén L, Jennings C, EUROASPIRE Investigators. EUROASPIRE IV: A European Society of Cardiology survey on the lifestyle, risk factor and therapeutic management of coronary patients from 24 European countries. Eur J Prev Cardiol 2016 Apr;23(6):636-648. [doi: 10.1177/2047487315569401] [Medline: 25687109]
4. Charlson M, Wells M, Peterson J, Boutin-Foster C, Ogedegbe G, Mancuso C, et al. Mediators and moderators of behavior change in patients with chronic cardiopulmonary disease: the impact of positive affect and self-affirmation. Transl Behav Med 2014 Mar;4(1):7-17 [FREE Full text] [doi: 10.1007/s13142-013-0241-0] [Medline: 24653772]
5. Habibović M, Burg M, Pedersen S. Behavioral interventions in patients with an implantable cardioverter defibrillator: lessons learned and where to go from here? Pacing Clin Electrophysiol 2013 May;36(5):578-590 [FREE Full text] [doi: 10.1111/pace.12108] [Medline: 23438053]
6. Minich DM, Bland JS. Personalized lifestyle medicine: relevance for nutrition and lifestyle recommendations. ScientificWorldJournal 2013;2013:129841 [FREE Full text] [doi: 10.1155/2013/129841] [Medline: 23878520]
7. Young L, Buse J, Weaver M, Vu M, Mitchell C, Blakeney T, Monitor Trial Group. Glucose self-monitoring in non-insulin-treated patients with type 2 diabetes in primary care settings: a randomized trial. JAMA Intern Med 2017 Jul 1;177(7):920-929 [FREE Full text] [doi: 10.1001/jamainternmed.2017.1233] [Medline: 28600913]
8. Kraal JJ, van den Akker-van Marle M, Abu-Hanna A, Stut W, Peek N, Kemps HM. Clinical and cost-effectiveness of home-based cardiac rehabilitation compared to conventional, centre-based cardiac rehabilitation: Results of the FIT@Home study. Eur J Prev Cardiol 2017 Aug;24(12):1260-1273 [FREE Full text] [doi: 10.1177/2047487317710803] [Medline: 28534417]
9. Sankaran S, Dendale P, Coninx K. Evaluating the impact of the hearthab app on motivation, physical activity, quality of life, and risk factors of coronary artery disease patients: Multidisciplinary crossover study. JMIR Mhealth Uhealth 2019 Apr 4;7(4):e10874 [FREE Full text] [doi: 10.2196/10874] [Medline: 30946021]

10. Kangovi S, Asch DA. Behavioral phenotyping in health promotion: embracing or avoiding failure. J Am Med Assoc 2018 May 22;319(20):2075-2076 [FREE Full text] [doi: 10.1001/jama.2018.2921] [Medline: 29710244]
11. Habibović M, Broers E, Piera-Jimenez J, Wetzels M, Ayoola I, Denollet J, et al. Enhancing lifestyle change in cardiac patients through the do CHANGE system ('Do cardiac health: advanced new generation ecosystem'): randomized controlled trial protocol. JMIR Res Protoc 2018 Feb 8;7(2):e40 [FREE Full text] [doi: 10.2196/resprot.8406] [Medline: 29422454]
12. Pine K, Fletcher B. Time to shift brain channels to bring about effective changes in health behaviour. Perspect Public Health 2014 Jan;134(1):16-17. [doi: 10.1177/1757913913514705] [Medline: 24395840]
13. Fletcher B, Hanson J, Page N, Pine K. FIT – Do something different. A new behavioral program for sustained weight loss. Swiss J Psychol 2011;70(1):25-34. [doi: 10.1024/1421-0185/a000035]
14. Walker SN, Sechrist KR, Pender NJ. Deep Blue. 1995. Health Promotion Model - Instruments to Measure Health Promoting Lifestyle : Health-Promoting Lifestyle Profile [HPLP II] URL: https://deepblue.lib.umich.edu/handle/2027.42/85349 [accessed 2018-06-30]
15. -. Development of the World Health Organization WHOQOL-BREF quality of life assessment. The WHOQOL Group. Psychol Med 1998 May;28(3):551-558. [doi: 10.1017/s0033291798006667] [Medline: 9626712]
16. Venkatesh V, Thong JY, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. Manag Inf Syst Q 2012;36(1):157-178. [doi: 10.2307/41410412]
17. Larsen DL, Attkisson C, Hargreaves WA, Nguyen TD. Assessment of client/patient satisfaction: development of a general scale. Eval Program Plann 1979;2(3):197-207. [doi: 10.1016/0149-7189(79)90094-6] [Medline: 10245370]
18. Denollet J. DS14: standard assessment of negative affectivity, social inhibition, and Type D personality. Psychosom Med 2005;67(1):89-97. [doi: 10.1097/01.psy.0000149256.81953.49] [Medline: 15673629]
19. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
20. Kroenke K, Spitzer R, Williams J. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
21. Stafford L, Berk M, Jackson HJ. Validity of the Hospital Anxiety and Depression Scale and Patient Health Questionnaire-9 to screen for depression in patients with coronary artery disease. Gen Hosp Psychiatry 2007;29(5):417-424. [doi: 10.1016/j.genhosppsych.2007.06.005] [Medline: 17888808]
22. Kelders SM, Kok RN, Ossebaard HC, van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
23. Alm-Roijer C, Fridlund B, Stagmo M, Erhardt L. Knowing your risk factors for coronary heart disease improves adherence to advice on lifestyle changes and medication. J Cardiovasc Nurs 2006;21(5):E24-E31. [doi: 10.1097/00005082-200609000-00015] [Medline: 16966907]
24. Riegel B, Moser D, Buck H, Dickson V, Dunbar S, Lee C, American Heart Association Council on Cardiovascular and Stroke Nursing; Council on Peripheral Vascular Disease;Council on Quality of Care and Outcomes Research. Self-care for the prevention and management of cardiovascular disease and stroke: a scientific statement for healthcare professionals from the American Heart Association. J Am Heart Assoc 2017 Aug 31;6(9):pii: e006997 [FREE Full text] [doi: 10.1161/JAHA.117.006997] [Medline: 28860232]
25. Habibović M, Denollet J, Cuijpers P, van der Voort PH, Herrman J, Bouwels L, et al. Web-based distress management for implantable cardioverter defibrillator patients: a randomized controlled trial. Health Psychol 2017 Apr;36(4):392-401. [doi: 10.1037/hea0000451] [Medline: 28192003]
26. Kashdan TB, Rottenberg J. Psychological flexibility as a fundamental aspect of health. Clin Psychol Rev 2010 Nov;30(7):865-878 [FREE Full text] [doi: 10.1016/j.cpr.2010.03.001] [Medline: 21151705]
27. Huffman JC, Celano CM, Beach SR, Motiwala SR, Januzzi JL. Depression and cardiac disease: epidemiology, mechanisms, and diagnosis. Cardiovasc Psychiatry Neurol 2013;2013:695925 [FREE Full text] [doi: 10.1155/2013/695925] [Medline: 23653854]
28. Lichtman JH, Froelicher ES, Blumenthal JA, Carney RM, Doering LV, Frasure-Smith N, American Heart Association Statistics Committee of the Council on Epidemiology and Prevention and the Council on Cardiovascular and Stroke Nursing. Depression as a risk factor for poor prognosis among patients with acute coronary syndrome: systematic review and recommendations: a scientific statement from the American Heart Association. Circulation 2014 Mar 25;129(12):1350-1369. [doi: 10.1161/CIR.0000000000000019] [Medline: 24566200]
29. Freedland KE, Rich MW, Skala JA, Carney RM, Dávila-Román VG, Jaffe AS. Prevalence of depression in hospitalized patients with congestive heart failure. Psychosom Med 2003;65(1):119-128. [doi: 10.1097/01.psy.0000038938.67401.85] [Medline: 12554823]
30. Sin NL, Kumar AD, Gehi AK, Whooley MA. Direction of association between depressive symptoms and lifestyle behaviors in patients with coronary heart disease: the heart and soul study. Ann Behav Med 2016 Aug;50(4):523-532 [FREE Full text] [doi: 10.1007/s12160-016-9777-9] [Medline: 26817654]
31. Penninx BW. Depression and cardiovascular disease: Epidemiological evidence on their linking mechanisms. Neurosci Biobehav Rev 2017 Mar;74(Pt B):277-286. [doi: 10.1016/j.neubiorev.2016.07.003] [Medline: 27461915]

Abbreviations

CAD: coronary artery disease
CAU: care as usual
Do CHANGE: Do Cardiac Health Advanced New Generation Ecosystem
DSD: Do Something Different
eHealth: electronic health
HF: heart failure
HT: hypertension
NA: negative affectivity
PHQ-9: Patient Health Questionnaire 9
SI: social inhibition
UTAUT2: Unified Theory of Acceptance 2
WHOQOL-BREF: World Health Organization Quality of Life—BREF

Edited by G Eysenbach; submitted 02.05.19; peer-reviewed by A Evers, C Loum, S Sankaran; comments to author 14.09.19; revised version received 30.10.19; accepted 15.12.19; published 22.05.20.

Please cite as:
Broers ER, Kop WJ, Denollet J, Widdershoven J, Wetzels M, Ayoola I, Piera-Jimenez J, Habibovic M
A Personalized eHealth Intervention for Lifestyle Changes in Patients With Cardiovascular Disease: Randomized Controlled Trial
J Med Internet Res 2020;22(5):e14570
URL: https://www.jmir.org/2020/5/e14570
doi:10.2196/14570
PMID:32441658

Original Paper

Occupational Health Needs and Predicted Well-Being in Office Workers Undergoing Web-Based Health Promotion Training: Cross-Sectional Study

Devan Richard Tchir[1], BSc, MPH; Michael Lorne Szafron[1], PhD

School of Public Health, University of Saskatchewan, Saskatoon, SK, Canada

Corresponding Author:
Michael Lorne Szafron, PhD
School of Public Health
University of Saskatchewan
104 Clinic Place
Saskatoon, SK, S7N 2Z4
Canada
Phone: 1 3069665468
Fax: 1 3069662264
Email: michael.szafron@usask.ca

Abstract

Background: Office workers face workplace-related health issues, including stress and back pain, resulting in considerable cost to businesses and health care systems. Workplace health promotion attempts to prevent these health issues, and the internet can be used to deliver workplace health promotion interventions to office workers. Data were provided by Fitbase GmbH, a German company, which specializes in workplace health promotion via the internet (Web-based health). The Web-based health intervention allowed workers to focus on different health categories by using information modules (reading health information) and/or completing practical exercises (guided, interactive health tutorials).

Objective: This study aimed to identify the extent to which office workers have workplace-related health issues, assess whether office workers who differ in their health focus also differ in their improved well-being, and assess whether completing practical exercises is associated with improved well-being compared with reading information modules.

Methods: Fitbase GmbH collected data for the period of February 2016 to May 2017 from health insurance employees undergoing Web-based health training in Hamburg, Germany. The data consisted of a needs assessment examining health issues faced by office workers, a wellness questionnaire regarding one's perception of the Web-based health intervention, and activity logs of information modules and practical exercises completed. Through logistic regression, we determined associations between improved well-being from Web-based health training and differences in a worker's health focus and a worker's preferred intervention method.

Results: Nearly half of the office workers had chronic back pain (1532/3354) and felt tense or irritated (1680/3348). Over four-fifth (645/766) of the office workers indicated that the Web-based health training improved their well-being (P<.001). Office workers who preferred practical exercises compared with information modules had 2.22 times greater odds of reporting improved well-being from the Web-based health intervention (P=.01; 95% CI 1.20-4.11). Office workers with a focus on practical exercises for back health had higher odds of improved well-being compared with other health foci. Office workers focused on practical exercises for back pain had at least two times the odds of having their well-being improved from the Web-based health intervention compared with those focused on stress management (P<.001), mindfulness (P=.02), stress management/mindfulness (P=.005), and eye health (P=.003). No particular health focus was associated with improved well-being for the information modules.

Conclusions: Office workers frequently report having back pain and stress. A focus on Web-based health training via practical exercises and practical exercises for back health predict an improvement in office workers' reported well-being.

(J Med Internet Res 2020;22(5):e14093) doi:10.2196/14093

KEYWORDS

eHealth; health promotion; occupational health; well-being; internet; workplace

Introduction

Background

Office workers face a multitude of health issues, including back pain [1], stress [2,3], eye problems [4], and upper limb pain [1]. Internationally, costs from back pain [5] and stress in the workplace [6] are extensive. Eye problems decrease the productivity of office workers [7], and upper extremity injuries result in long absences from work and are a financial burden to businesses [8]. All of these health afflictions can have a significant financial toll through loss of work and increased health care expenses.

Workplace health promotion can be used to prevent many workplace-related health concerns and can include a range of information delivery systems. These systems include intervention programs, disease management programs, print materials, health education classes, fitness facilities on worksites, health fairs, and interventions for diet and fitness, among other methods [9]. Web-based electronic health interventions (Web-based health) are also available, which can be used to address the stated health issues. Web-based health interventions have been shown to decrease back pain [10], help manage stress [9,11], and provide improvements to nutritional and physical activity habits [9,12]. Addressing health issues via Web-based health techniques is favorable given that they can be personalized, highly structured, visually stimulating, and readily available to workers [13]. Furthermore, internet-based interventions have been shown to dramatically restore work capacity [14], have more favorable changes in health behavior and knowledge compared with non–Web-based interventions [15], and improve worker well-being [16].

Self-reported well-being and self-reported health have been beneficial predictors for decreased sick days, decreased costs to the health care system, and increased productivity [17-20]. There is also evidence that well-being is negatively associated with spinal pain [21], stress [22], eye problems [23], and mortality [24]. Thus, self-reported well-being as an outcome of Web-based health training can be an important measure. If a worker's reported well-being differs from his or her health concern (such as stress management or back pain) is an important question. Although there is literature documenting the effects of Web-based health interventions on worker well-being (through measures of stress, psychological distress, depression scores, etc) [16], we found little literature documenting the effect of Web-based health interventions on a worker directly reporting that their well-being had been improved from an intervention (defined in this work as *self-reported well-being*). Furthermore, many Web-based health interventions to our knowledge aim to improve a worker's psychological well-being through cognitive therapies and stress management [16], not via physical health improvement (ie, back health).

Web-based health promotion interventions primarily aim to change health behaviors through education and may involve the worker passively reading or listening to health information without engaging workers in practical, hands-on exercises designed to improve health. The effectiveness of practical, hands-on exercises (as opposed to simply reading information modules) in Web-based interventions is undocumented in the literature to the best of our knowledge. Also, information on self-perceived health issues (related to back pain, stress, etc) for office workers goes largely unreported.

Objectives

Our study was designed to address these two gaps in the literature. This study has the following research objectives: (1) identify the extent to which office workers have health issues related to repetitive strain injury (RSI, ie, upper limb pain associated with computer work), back pain, resilience, mindfulness, nutrition, stress, and eye health; (2) to assess if office workers who differ in their health focus also differ in their self-reported well-being; and (3) evaluate if completing practical exercises influences the self-reported well-being of office workers more than reading information modules.

Methods

Data

Data regarding health insurance employees in Hamburg, Germany, undergoing Web-based health training for the period of February 2016 to May 2017, was sourced from Fitbase GmbH, a German company based out of Hamburg, Germany, specializing in Web-based health training for companies and individuals. Their Web-based health training intervention consists of reading educational information modules (with health categories in back pain, stress management, nutrition, mindfulness, eye health, RSI, and resilience), and practical exercises (with health categories in back pain, stress management, mindfulness, eye health, and RSI). The information modules educate workers on maintaining a healthy lifestyle, and the practical exercises are guided health tutorials for activities related to healthy behavior. The data included responses to a needs assessment, a questionnaire, and practical exercise and information module logs of the employees. For privacy reasons, no employee demographic information was made available.

The total number of employees enrolled in the Web-based health training was 5694. The needs assessment was administered before beginning the Web-based health training and consisted of statements associated with RSI, back pain, resilience, mindfulness, nutrition, stress management, and eye health. These statements were, respectively: (1) "After intensive PC (personal computer) work, I feel pain, tingling or numbness in my hands or arms"; (2) "I have upper and lower back pain"; (3) "I basically assume that I can overcome difficulties in life"; (4) "I notice that I'm lost in thought about the future or the past"; (5) "I crave fast food"; (6) "I feel tense or irritated"; and (7) "My eyes hurt after work." Responses to these statements were *never*, *rarely*, *sometimes*, *frequently*, and *constantly*. Questionnaires regarding perceptions of the Web-based health training were administered during training at roughly 3, 6, and 9 weeks and were completed voluntarily. The last questionnaire completed was used for all analyses to best capture an individual's perception at the end of training. The statement of interest on the questionnaire was "Online training improves my well-being," answered by a worker as *strongly disagree*,

disagree, *agree*, or *strongly agree*. The University of Saskatchewan Research Ethics Board waived the need for ethics approval based on secondary analyses of the data.

Variables

Outcome Variable

Those who responded *strongly disagree* or *disagree* to the statement on the questionnaire "Online training improves my well-being" were coded as *disagree*, and those who responded *agree* or *strongly agree* were coded as *agree*, creating a binary outcome variable. These responses were combined due to a small sample size.

Independent Variables: Preferred Type of Intervention, Practical Exercise Health Focus, and Information Health Focus

The practical exercises completed and information modules read throughout the Web-based health training were recorded per user. On the basis of these logs, we computed the total numbers of practical exercises completed, and information modules read per health category. From these totals, we computed per individual the grand total of practical exercises completed and the grand total of the information modules read. We then defined a user's *preferred type of intervention* as the intervention associated with the greater of these two grand totals; if these grand totals were not different, the workers *preferred type of intervention* was set to be *no preference*. For example, if a worker completed statistically more practical exercises than information modules, they were assigned a *practical exercises* preference.

A user's *practical exercise health focus* was defined to be the practical exercise health category associated with the maximum total of the practical exercise-related health categories. For example, if a worker completed most of their practical exercises in stress management, they were assigned a *stress management* focus. A user's *information health focus* was derived similarly. If two health categories were completed most frequently, a tie was assigned to the worker. The number of ties was negligible (each combination <1%) except for workers completing both stress and mindfulness, and both back pain and RSI practical exercises. Consequently, we respectively labeled their *practical exercise health focus* as *stress and mindfulness* and *back and RSI*, and we excluded the data for individuals with all other ties.

Analyses

To address our first research objective (which was to identify the extent office workers experience health issues such as RSI, back pain, resilience, mindfulness, nutrition, stress, and eye health), we summarized the frequencies of the needs assessment data. From these frequencies we found the proportion of office workers who responded as *never*, *rarely*, *sometimes*, *frequently*, and *constantly* to the questions/statements: (1) "After intensive PC work, I feel pain, tingling or numbness in my hands or arms" (n=3354); (2) "I have upper and lower back pain" (n=3354); (3) "I basically assume that I can overcome difficulties in life" (n=3348); (4) "I notice that I'm lost in thought about the future or the past" (n=3354); (5) "I crave fast food" (n=3354); (6) "I feel tense or irritated" (n=3348); and (7) "My eyes hurt after work" (n=3348).

To address our second and third objectives, we developed a logistic regression model that predicted the well-being of workers from the Web-based health training. Those who completed at least two practical exercises, two information modules, and a questionnaire were included for the logistic regression analysis (n=779). The response to the statement, "Online training improves my well-being," was used as the outcome variable. Independent variables included the derived variables *practical exercise health focus*, *information health focus*, and *preferred type of intervention*. Office workers with an *RSI* categorized *information health focus* (n=7) and with a *back and RSI* categorized *practical exercise health focus* (n=6) were removed due to the extremely small sample size and were not included in the final model (n=766). Due to insufficient sample size, interactions between variables were not considered in the model.

Bivariate analyses of the independent variables with the outcome variable having $P<.20$ were to be included in the model, and all subsequent analyses were completed at the alpha=.05 level. All variance inflation factor scores were below 1.7, indicating a negligible influence of multicollinearity. All statistical analyses were conducted using SAS version 9.4.

Results

User Statistics

Of the 934 workers who completed a questionnaire, 765 (81.9%) indicated that the Web-based health training improved their well-being whereas 169 (18.1%) indicated the Web-based health training had not improved their well-being ($P<.001$).

Of the workers who completed two or more practical exercises, 29.69% (1057/3560) focused on back pain, 21.94% (781/3560) focused on stress management, 18.93% (674/3560) focused on mindfulness, 13.88% (494/3560) focused on stress and mindfulness equally, 10.81% (385/3560) focused on eye health, and 2.70% (96/3560) focused on RSI (Table 1).

Of the workers who completed two or more information modules, 46.07% (2415/5242) focused on back pain, 18.41% (965/5242) focused on stress management, 17.89% (938/5242) focused on nutrition, 7.34% (385/5242) focused on mindfulness, 5.32% (279/5242) focused on eye health, 2.61% (137/5242) focused on RSI, and 2.35% (123/5242) focused on resilience (Table 2). Among the workers who completed a questionnaire and did two or more information modules and practical exercises, 84.2% (645/766) reported that the Web-based training had improved their well-being ($P<.001$; Table 3).

Table 1. Number of office workers categorized in each practical exercise health focus who completed at least two practical exercises (n=3560), and the number of office workers categorized in each practical exercise health focus who completed at least two practical exercises, at least two information modules, and a questionnaire (n=779).

Health focus	Office workers categorized in each practical exercise health focus who completed practical exercises (N=3560), n (%)	Office workers categorized in each practical exercise health focus who completed information modules, practical exercises, and a questionnaire (N=779), n (%)
Back pain	1057 (29.69)	226 (29.0)
Stress management	781 (21.94)	213 (27.3)
Mindfulness	674 (18.93)	162 (20.8)
Stress and mindfulness	494 (13.88)	76 (9.8)
Eye health	385 (10.81)	73 (9.4)
RSI[a]	96 (2.70)	23 (3.0)
Back and RSI	73 (2.05)	6 (0.8)

[a]RSI: repetitive strain injury.

Table 2. Number of office workers categorized in each information health focus who completed at least two information modules (n=5242), and the number of office workers categorized in each information health focus who completed at least two information modules, at least two practical exercises, and a questionnaire (n=779).

Health focus	Office workers categorized in each information health focus who completed information modules (N=5242), n (%)	Office workers categorized in each information health focus who completed information modules, practical exercises, and a questionnaire (N=779), n (%)
Back pain	2415 (46.07)	345 (44.3)
Stress management	965 (18.41)	234 (30.0)
Nutrition	938 (17.89)	96 (12.3)
Mindfulness	385 (7.34)	51 (6.6)
Eye health	279 (5.32)	24 (3.1)
RSI[a]	137 (2.61)	7 (0.9)
Resilience	123 (2.35)	22 (2.8)

[a]RSI: repetitive strain injury.

Table 3. Descriptive statistics of the variables considered for the logistic regression model and bivariate logistic regression analyses of the independent variables with the outcome variable "online training improves my well-being" (n=766).

Independent variable	Number of workers (N=766), n (%)	*P* value
Practical exercise health focus		.003
Back pain	224 (29.2)	
Stress management	211 (27.6)	
Mindfulness	162 (21.2)	
Stress and mindfulness	76 (9.9)	
Eye health	72 (9.4)	
RSI[a]	21 (2.7)	
Information health focus		.12
Back pain	344 (44.9)	
Stress management	233 (30.4)	
Mindfulness	51 (6.7)	
Nutrition	92 (12.0)	
Eye health	24 (3.1)	
Resilience	22 (2.9)	
Preferred type of intervention		.002
Practical exercises	208 (27.2)	
Information modules	351 (45.8)	
No preference	207 (27.0)	
Worker reports that online training improves their well-being		N/A[b]
Strongly agree/agree	645 (84.2)	
Strongly disagree/disagree	121 (15.8)	

[a]RSI: repetitive strain injury.

[b]N/A: not applicable.

Research Objective 1

Our first research objective was to identify the extent to which office workers have health issues related to different health categories. Our results related to this are the distributions of the responses to the needs assessment questions as illustrated in Table 4. Over 42.60% (1429/3354) never felt pain, tingling, or numbness in their hands or arms after intensive PC work. Nearly half of the office workers (1532/3354) chronically had back pain. About three-fourth (2554/3348) believed they can consistently overcome difficulties in life. Half of the office workers (1680/3348) felt chronically tense or irritated. About one-fifth of office workers (643/3348) had eye pain after work frequently or more often.

Research Objectives 2 and 3

The results of the logistic regression model, including *P* values, odds ratios (ORs), and 95% CIs, could be found in Table 5. For our second research objective, which was to assess if office workers who differ in their health focus differ in their self-reported well-being, we found one's *practical exercise health focus* to be statistically significant (Wald chi-square value, χ^2_5=14.5; *P*=.01) with a worker's improved well-being from the Web-based health training. For our third research objective, which was to evaluate if completing practical exercises influence self-reported well-being more than information modules, we found one's *preferred type of intervention* to also be statistically significant (Wald chi-square value, χ^2_2=6.7; *P*=.03) with a worker's improved well-being from the Web-based health training.

Table 4. Needs assessment frequency data of office workers in Germany. Number (%) represents the office workers and their responses to needs assessment statements (n=3354).

Needs assessment statements	Never, n (%)	Rarely, n (%)	Sometimes, n (%)	Frequently, n (%)	Constantly, n (%)	Missing, n (%)
"I basically assume I can overcome difficulties in life."	15 (0.45)	147 (4.38)	632 (18.84)	1655 (49.34)	899 (26.80)	6 (0.18)
"I notice that I'm lost in thought about the future or the past."	392 (11.69)	936 (27.91)	1025 (30.56)	787 (23.46)	214 (6.38)	0 (0.00)
"After intensive PC work, I feel pain, tingling or numbness in my hands or arms."	1429 (42.61)	880 (26.24)	671 (20.01)	325 (9.69)	49 (1.46)	0 (0.00)
"I have upper and lower back pain."	145 (4.32)	618 (18.43)	1059 (31.57)	1080 (32.20)	452 (13.48)	0 (0.00)
"I crave fast food."	180 (5.37)	768 (22.90)	1115 (33.24)	967 (28.83)	324 (9.66)	0 (0.00)
"I feel tense or irritated."	54 (1.61)	399 (11.90)	1215 (36.23)	1302 (38.82)	378 (11.27)	6 (0.18)
"My eyes hurt after work."	747 (22.27)	1003 (29.90)	955 (28.47)	538 (16.04)	105 (3.13)	6 (0.18)

Table 5. The associations of different health foci and interventions with the well-being of office workers in Germany undergoing Web-based health training (n=766).

Independent variables	OR[a] (95% CI)	1/OR[b]	*P* value
Practical exercise health focus			.01[c]
Back pain (reference category)	N/A[d]	N/A	N/A
Stress management	0.322 (0.166-0.623)	3.109	<.001
Mindfulness	0.439 (0.215-0.895)	2.278	.02
Stress/mindfulness	0.319 (0.143-0.710)	3.137	.005
Eye health	0.291 (0.131-0.648)	3.436	.003
RSI[e]	0.548 (0.142-2.124)	1.824	.38
Preferred type of intervention			.03[c]
Information modules (reference category)	N/A	N/A	N/A
Practical exercises	2.218 (1.196-4.112)	0.451	.01
No preference	1.083 (0.671-1.749)	0.923	.74
Information health focus			.37[c]
Back pain (reference category)	N/A	N/A	N/A
Stress management	1.097 (0.652-1.848)	0.911	.73
Mindfulness	0.704 (0.295-1.684)	1.42	.43
Nutrition	0.643 (0.354-1.168)	1.556	.15
Eye health	1.574 (0.411-6.029)	0.635	.51
Resilience	1.979 (0.419-9.359)	0.505	.39

[a]OR: odds ratio.

[b]1/OR: reciprocal of the odds ratio.

[c]*P* value estimated from a Type III Analysis of Effects via a Wald χ^2 test.

[d]N/A: not applicable.

[e]RSI: repetitive strain injury.

Compared with workers focused on practical exercises for back pain, workers focused on eye health, mindfulness, stress management, and stress/mindfulness statistically differed (refer to Table 5 for the respective *P* values). Those workers with a focus on practical exercises for back pain had 3.11 times the odds greater than those focused on stress management (*P*<.001), 2.28 times the odds greater than those focused on mindfulness (*P*=.02), 3.14 times the odds greater than those focused on stress/mindfulness (*P*=.005), and 3.44 times the odds greater than those focused in eye health practical exercises (*P*=.003) of having their well-being improved.

Among office workers whose preferred intervention was practical exercises, the odds were 2.22 times greater to report improved well-being from the Web-based health training than those who preferred information modules (*P*=.01).

Discussion

Principal Findings

This study sought to examine the level of importance in which office workers perceived health issues, and examined possible differences in improved well-being from Web-based health training. We were interested in if there are differences in the associations for a worker's well-being based on specific health categories and if completing guided practical exercises or reading educational information modules had a different impact on improving an office worker's well-being.

In our exploration of office worker work-related health issues, back pain and stress management were most important; mindfulness and nutrition were moderately important; and RSI, resiliency, and eye health were of least concern to office workers. Back pain and stressful demands from work are common among office workers [25]. This aligns with the popularity of these two health categories in our study, as they were the most popular practical exercises and information modules completed by workers. Studies have reported a wide range of back pain prevalence (ranging from 23% to 56%) in office workers [1,26-29]. This is consistent with our finding that workers frequently focused on back pain when completing both practical exercises and information modules (29.69% and 46.07%, respectively).

Office workers may suffer from stress [2,3], and our study cohort was no different. Half of the office workers in this study felt chronically tense or irritated. However, over three-quarters of office workers in our study stated that they frequently or constantly could overcome life difficulties, suggesting our study participants are intrinsically confident in their abilities to overcome stressors. This indicates an apparent dichotomy between a worker experiencing stress and their resilience.

Oha et al [27] found that 7% of computer workers reported frequent wrist/hand or shoulder pain. Other studies report varying results; the prevalence of upper limb symptoms as low as 10% and as high as 52% have been documented in office workers [1,27-30]. Our findings support that workers experience varying degrees of pain after PC work, whereas most studies simply report if workers experience upper limb pain in a dichotomous fashion [1,28-31]. In fact, those responding *constantly feeling pain after PC work* might have disabling pain or worse [32]. In our study, 11.15% of workers chronically felt pain in their hands or arms after PC work, compared with 68.84% of workers reporting rarely or never. These estimates are considerably different from the literature, which estimates a higher prevalence of RSI pain [29]. Only 19.21% of our office workers reported routinely feeling eye pain after work. Another study identifying health issues in data processing office workers report a 26% prevalence of sore eyes from work [28].

Our logistic regression model yielded that well-being was influenced by different health categories. Workers focused on practical exercises for back pain had higher odds of having their well-being improved compared with almost all health categories. Prior literature has demonstrated that Web-based back pain interventions for office workers improves their quality of life [10], possibly consistent with our findings. Our logistic regression model also yielded that well-being was influenced by if office workers completed more practical exercises or more information modules. In general, completing practical exercises improved a worker's odds of having their well-being improved. Our findings suggest that completing practical exercises can be more effective in improving well-being as opposed to passively reading information modules, thus suggesting a need for hands-on, instructional, do-it-yourself exercises in Web-based health promotion for the office workplace.

Furthermore, participation rates in Web-based health promotion activities are typically low and hence a major concern [33]. Evidence suggests that guided Web-based health programs promote worker retention [16]. Our finding that office workers have increased odds of improved well-being when utilizing guided, follow-along exercises might reflect this. Available literature shows exercise-driven back pain interventions can decrease pain and improve back function [34,35]. Our study suggests back health might be in greatest demand by office workers; this makes sense, given that many different occupations are linked with back pain [36]. For groups with limited resources, their worker's health might benefit the most from focusing on back health.

Overall, an emphasis on practical exercises for back pain might benefit a workplace to the greatest extent when focusing on a worker population's well-being. For our office worker population, we found that back pain was the most popular health focus and was perceived as a persistent health issue (although those who focused on practical exercises for back pain had higher odds of having their well-being improved). It is possible that different working populations have different health needs. Our work may suggest that workers should address their primary health concern with practical exercises for maximal improvement in their perceived well-being.

To the best of our knowledge, our findings are the first demonstrating that completing Web-based health training is associated with improved self-reported well-being in office workers and that completing practical exercises rather than simply receiving Web-based health information may improve well-being. Our work supports that practical exercises for Web-based health interventions are beneficial to Web-based health promotion as a whole, regardless of a specific health

focus, in office workers. With respect to Web-based health promotion in the workplace, our work suggests that actively engaging workers to take part in their health is more beneficial than passively engaging workers. The data suggests that Web-based health interventions for an office-workplace should tailor resources toward promoting back health and in general, exercise-intensive health promotion strategies.

Given the potential cost benefits of reported well-being, employers and health insurers could benefit from incorporating practical back health exercises into Web-based health promotion efforts. As there is a shift in employment to many sedentary and office-confined occupations [37] and the internet is widely used globally [38], our findings might be applicable to a wider spectrum of occupations.

Limitations

As only those who completed a questionnaire, practical exercises, and information modules were included in our logistic regression modeling, there is potential selection bias. The absence of demographic information on the office workers prevented us from controlling for socio-demographic confounding factors. The findings were centered on German office workers and therefore may not generalize to other countries. Health categories, from which workers chose, were not exhaustive; there are likely other important concerns that were not included in this study, such as physical activity [39]. Future studies should identify objective measures (blood pressure, cortisol levels, etc) and a way to compare the perceived quality of the practical exercises to information modules.

Conclusions

Our finding suggests that a Web-based health promotion intervention using hands-on activities can be an effective method for addressing workplace-related health concerns and general worker well-being. We recommend future Web-based health promotion workplace interventions should focus on the main health issues experienced by workers and to include guided practical exercises for those who are wanting to improve their health. As technology becomes more mainstream in the workforce and jobs continue to shift toward seated computer work, Web-based health promotion will become increasingly more important.

Acknowledgments

The authors thank Fitbase GmbH and all of their employees for collecting and making data available, as well as the DAAD Research Internships in Science and Engineering professional program and the Dr James Rossiter Master's of Public Health Practicum Award for financial support.

Conflicts of Interest

DT was a former intern at Fitbase GmbH, but this work was completed after DT's internship, and DT has no financial incentives. Fitbase had no role in the design, practice, or analysis of this study.

References

1. Janwantanakul P, Pensri P, Jiamjarasrangsri V, Sinsongsook T. Prevalence of self-reported musculoskeletal symptoms among office workers. Occup Med (Lond) 2008 Sep;58(6):436-438. [doi: 10.1093/occmed/kqn072] [Medline: 18544589]
2. Krantz G, Berntsson L, Lundberg U. Total workload, work stress and perceived symptoms in Swedish male and female white-collar employees. Eur J Public Health 2005 Apr;15(2):209-214. [doi: 10.1093/eurpub/cki079] [Medline: 15941764]
3. Tsai SY. A study of the health-related quality of life and work-related stress of white-collar migrant workers. Int J Environ Res Public Health 2012 Oct 19;9(10):3740-3754 [FREE Full text] [doi: 10.3390/ijerph9103740] [Medline: 23202771]
4. Portello JK, Rosenfield M, Bababekova Y, Estrada JM, Leon A. Computer-related visual symptoms in office workers. Ophthalmic Physiol Opt 2012 Sep;32(5):375-382. [doi: 10.1111/j.1475-1313.2012.00925.x] [Medline: 22775070]
5. Dagenais S, Caro J, Haldeman S. A systematic review of low back pain cost of illness studies in the United States and internationally. Spine J 2008;8(1):8-20. [doi: 10.1016/j.spinee.2007.10.005] [Medline: 18164449]
6. Hassard J, van den Broek K, Flemming D, Gründler R, Dewe P, Teoh K, et al. Calculating the Cost of Work-Related Stress and Psychosocial Risks. Luxembourg: European Agency for Safety and Health at Work; 2014.
7. Yamada M, Mizuno Y, Shigeyasu C. Impact of dry eye on work productivity. Clinicoecon Outcomes Res 2012;4:307-312 [FREE Full text] [doi: 10.2147/CEOR.S36352] [Medline: 23091391]
8. Hashemi L, Webster BS, Clancy EA, Courtney TK. Length of disability and cost of work-related musculoskeletal disorders of the upper extremity. J Occup Environ Med 1998 Mar;40(3):261-269. [doi: 10.1097/00043764-199803000-00008] [Medline: 9531097]
9. Cook RF, Billings DW, Hersch RK, Back AS, Hendrickson A. A field test of a web-based workplace health promotion program to improve dietary practices, reduce stress, and increase physical activity: randomized controlled trial. J Med Internet Res 2007 Jun 19;9(2):e17 [FREE Full text] [doi: 10.2196/jmir.9.2.e17] [Medline: 17581811]
10. Del Pozo-Cruz B, Adsuar JC, Parraca J, Del Pozo-Cruz J, Moreno A, Gusi N. A web-based intervention to improve and prevent low back pain among office workers: a randomized controlled trial. J Orthop Sports Phys Ther 2012 Oct;42(10):831-841. [doi: 10.2519/jospt.2012.3980] [Medline: 22951407]

11. Heber E, Ebert DD, Lehr D, Cuijpers P, Berking M, Nobis S, et al. The benefit of web- and computer-based interventions for stress: a systematic review and meta-analysis. J Med Internet Res 2017 Feb 17;19(2):e32 [FREE Full text] [doi: 10.2196/jmir.5774] [Medline: 28213341]
12. Storm V, Dörenkämper J, Reinwand DA, Wienert J, de Vries H, Lippke S. Effectiveness of a web-based computer-tailored multiple-lifestyle intervention for people interested in reducing their cardiovascular risk: a randomized controlled trial. J Med Internet Res 2016 Apr 11;18(4):e78 [FREE Full text] [doi: 10.2196/jmir.5147] [Medline: 27068880]
13. Ritterband LM, Gonder-Frederick LA, Cox DJ, Clifton AD, West RW, Borowitz SM. Internet interventions: in review, in use, and into the future. Prof Psychol Res Pr 2003;34(5):527-534. [doi: 10.1037/0735-7028.34.5.527]
14. Brattberg G. Internet-based rehabilitation for individuals with chronic pain and burnout II: a long-term follow-up. Int J Rehabil Res 2007 Sep;30(3):231-234. [doi: 10.1097/MRR.0b013e32829fa545] [Medline: 17762769]
15. 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]
16. Carolan S, Harris PR, Cavanagh K. Improving employee well-being and effectiveness: systematic review and meta-analysis of web-based psychological interventions delivered in the workplace. J Med Internet Res 2017 Jul 26;19(7):e271 [FREE Full text] [doi: 10.2196/jmir.7583] [Medline: 28747293]
17. Halford C, Wallman T, Welin L, Rosengren A, Bardel A, Johansson S, et al. Effects of self-rated health on sick leave, disability pension, hospital admissions and mortality. A population-based longitudinal study of nearly 15,000 observations among Swedish women and men. BMC Public Health 2012 Dec 22;12:1103 [FREE Full text] [doi: 10.1186/1471-2458-12-1103] [Medline: 23259777]
18. Hornquist J, Hansson B, Zar M. Well-being and future sick-leave: multivariate analyses with regard to preceding sick-leave. Eur J Public Health 1997 Sep 1;7(3):284-290. [doi: 10.1093/eurpub/7.3.284]
19. George JM. Mood and absence. J Appl Psychol 1989 Apr;74(2):317-324. [doi: 10.1037/0021-9010.74.2.317]
20. Danna K, Griffin RW. Health and well-being in the workplace: A review and synthesis of the literature. J Manage 1999;25(3):357-384. [doi: 10.1177/014920639902500305]
21. Stallknecht SE, Strandberg-Larsen K, Hestbæk L, Andersen AN. Spinal pain and co-occurrence with stress and general well-being among young adolescents: a study within the Danish National Birth Cohort. Eur J Pediatr 2017 Jun;176(6):807-814. [doi: 10.1007/s00431-017-2915-y] [Medline: 28470441]
22. Moulin F, Keyes C, Liu A, Caron J. Correlates and predictors of well-being in Montreal. Community Ment Health J 2017 Jul;53(5):560-567. [doi: 10.1007/s10597-017-0080-4] [Medline: 28160126]
23. Lee PP, Spritzer K, Hays RD. The impact of blurred vision on functioning and well-being. Ophthalmology 1997 Mar;104(3):390-396. [doi: 10.1016/s0161-6420(97)30303-0] [Medline: 9082261]
24. Chida Y, Steptoe A. Positive psychological well-being and mortality: a quantitative review of prospective observational studies. Psychosom Med 2008 Sep;70(7):741-756. [doi: 10.1097/PSY.0b013e31818105ba] [Medline: 18725425]
25. Schreuder KJ, Roelen CA, Koopmans PC, Groothoff JW. Job demands and health complaints in white and blue collar workers. Work 2008;31(4):425-432. [Medline: 19127013]
26. Juul-Kristensen B, Søgaard K, Strøyer J, Jensen C. Computer users' risk factors for developing shoulder, elbow and back symptoms. Scand J Work Environ Health 2004 Oct;30(5):390-398 [FREE Full text] [doi: 10.5271/sjweh.827] [Medline: 15529802]
27. Oha K, Animägi L, Pääsuke M, Coggon D, Merisalu E. Individual and work-related risk factors for musculoskeletal pain: a cross-sectional study among Estonian computer users. BMC Musculoskelet Disord 2014 May 28;15:181 [FREE Full text] [doi: 10.1186/1471-2474-15-181] [Medline: 24884911]
28. Woods V. Musculoskeletal disorders and visual strain in intensive data processing workers. Occup Med (Lond) 2005 Mar;55(2):121-127. [doi: 10.1093/occmed/kqi029] [Medline: 15757986]
29. Kaliniene G, Ustinaviciene R, Skemiene L, Vaiciulis V, Vasilavicius P. Associations between musculoskeletal pain and work-related factors among public service sector computer workers in Kaunas County, Lithuania. BMC Musculoskelet Disord 2016 Oct 7;17(1):420 [FREE Full text] [doi: 10.1186/s12891-016-1281-7] [Medline: 27717347]
30. van den Heuvel SG, van der Beek AJ, Blatter BM, Bongers PM. Do work-related physical factors predict neck and upper limb symptoms in office workers? Int Arch Occup Environ Health 2006 Aug;79(7):585-592. [doi: 10.1007/s00420-006-0093-8] [Medline: 16710709]
31. Kuorinka I, Jonsson B, Kilbom A, Vinterberg H, Biering-Sørensen F, Andersson G, et al. Standardised Nordic questionnaires for the analysis of musculoskeletal symptoms. Appl Ergon 1987 Sep;18(3):233-237. [doi: 10.1016/0003-6870(87)90010-x] [Medline: 15676628]
32. Coggon D, Ntani G, Palmer KT, Felli VE, Harari R, Barrero LH, et al. Disabling musculoskeletal pain in working populations: is it the job, the person, or the culture? Pain 2013 Jun;154(6):856-863 [FREE Full text] [doi: 10.1016/j.pain.2013.02.008] [Medline: 23688828]
33. Jimenez P, Bregenzer A. Integration of eHealth tools in the process of workplace health promotion: proposal for design and implementation. J Med Internet Res 2018 Feb 23;20(2):e65 [FREE Full text] [doi: 10.2196/jmir.8769] [Medline: 29475828]

34. Hayden JA, van Tulder MW, Tomlinson G. Systematic review: strategies for using exercise therapy to improve outcomes in chronic low back pain. Ann Intern Med 2005 May 3;142(9):776-785. [doi: 10.7326/0003-4819-142-9-200505030-00014] [Medline: 15867410]
35. Lidegaard M, Jensen RB, Andersen CH, Zebis MK, Colado JC, Wang Y, et al. Effect of brief daily resistance training on occupational neck/shoulder muscle activity in office workers with chronic pain: randomized controlled trial. Biomed Res Int 2013;2013:262386 [FREE Full text] [doi: 10.1155/2013/262386] [Medline: 24490152]
36. Punnett L, Prüss-Utün A, Nelson DI, Fingerhut MA, Leigh J, Tak S, et al. Estimating the global burden of low back pain attributable to combined occupational exposures. Am J Ind Med 2005 Dec;48(6):459-469. [doi: 10.1002/ajim.20232] [Medline: 16299708]
37. Church TS, Thomas DM, Tudor-Locke C, Katzmarzyk PT, Earnest CP, Rodarte RQ, et al. Trends over 5 decades in US occupation-related physical activity and their associations with obesity. PLoS One 2011;6(5):e19657 [FREE Full text] [doi: 10.1371/journal.pone.0019657] [Medline: 21647427]
38. Poushter J. Pew Research Center. 2016 Feb 22. Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies URL: https://tinyurl.com/yd3ph98l [accessed 2019-03-21] [WebCite Cache ID 772v9L8Xm]
39. Malik SH, Blake H, Suggs LS. A systematic review of workplace health promotion interventions for increasing physical activity. Br J Health Psychol 2014 Mar;19(1):149-180. [doi: 10.1111/bjhp.12052] [Medline: 23827053]

Abbreviations

OR: odds ratio
PC: personal computer
RSI: repetitive strain injury

Edited by G Eysenbach; submitted 05.04.19; peer-reviewed by P Jimenez, JR Bautista, Y Albalawi; comments to author 03.10.19; revised version received 28.11.19; accepted 28.02.20; published 26.05.20.

Please cite as:
Tchir DR, Szafron ML
Occupational Health Needs and Predicted Well-Being in Office Workers Undergoing Web-Based Health Promotion Training: Cross-Sectional Study
J Med Internet Res 2020;22(5):e14093
URL: http://www.jmir.org/2020/5/e14093/
doi:10.2196/14093
PMID:32452806

Original Paper

Integrating People, Context, and Technology in the Implementation of a Web-Based Intervention in Forensic Mental Health Care: Mixed-Methods Study

Hanneke Kip[1,2], MSc; Floor Sieverink[1], PhD; Lisette J E W C van Gemert-Pijnen[1,3], PhD; Yvonne H A Bouman[2], PhD; Saskia M Kelders[1,4], PhD

[1]Centre for eHealth and Wellbeing Research, Department of Psychology, Health and Technology, University of Twente, Enschede, Netherlands

[2]Department of Research, Transfore, Deventer, Netherlands

[3]Faculty of Medical Sciences, Universitair Medisch Centrum Groningen, Groningen, Netherlands

[4]Optentia Research Focus Area, North-West University, Vanderbijlpark, South Africa

Corresponding Author:
Hanneke Kip, MSc
Centre for eHealth and Wellbeing Research
Department of Psychology, Health and Technology
University of Twente
Drienerlolaan 5
Enschede
Netherlands
Phone: 31 053 489 9111
Email: h.kip@utwente.nl

Abstract

Background: While eMental health interventions can have many potential benefits for mental health care, implementation outcomes are often disappointing. In order to improve these outcomes, there is a need for a better understanding of complex, dynamic interactions between a broad range of implementation-related factors. These interactions and processes should be studied holistically, paying attention to factors related to context, technology, and people.

Objective: The main objective of this mixed-method study was to holistically evaluate the implementation strategies and outcomes of an eMental health intervention in an organization for forensic mental health care.

Methods: First, desk research was performed on 18 documents on the implementation process. Second, the intervention's use by 721 patients and 172 therapists was analyzed via log data. Third, semistructured interviews were conducted with all 18 therapists of one outpatient clinic to identify broad factors that influence implementation outcomes. The interviews were analyzed via a combination of deductive analysis using the nonadoption, abandonment, scale-up, spread, and sustainability framework and inductive, open coding.

Results: The timeline generated via desk research showed that implementation strategies focused on technical skills training of therapists. Log data analyses demonstrated that 1019 modules were started, and 18.65% (721/3865) of patients of the forensic hospital started at least one module. Of these patients, 18.0% (130/721) completed at least one module. Of the therapists using the module, 54.1% (93/172 sent at least one feedback message to a patient. The median number of feedback messages sent per therapist was 1, with a minimum of 0 and a maximum of 460. Interviews showed that therapists did not always introduce the intervention to patients and using the intervention was not part of their daily routine. Also, therapists indicated patients often did not have the required conscientiousness and literacy levels. Furthermore, they had mixed opinions about the design of the intervention. Important organization-related factors were the need for more support and better integration in organizational structures. Finally, therapists stated that despite its current low use, the intervention had the potential to improve the quality of treatment.

Conclusions: Synthesis of different types of data showed that implementation outcomes were mostly disappointing. Implementation strategies focused on technical training of therapists, while little attention was paid to changes in the organization, design of the technology, and patient awareness. A more holistic approach toward implementation strategies—with more attention to the organization, patients, technology, and training therapists—might have resulted in better implementation outcomes. Overall, adaptivity appears to be an important concept in eHealth implementation: a technology should be easily adaptable to an individual

patient, therapists should be trained to deal flexibly with an eMental health intervention in their treatment, and organizations should adapt their implementation strategies and structures to embed a new eHealth intervention.

(J Med Internet Res 2020;22(5):e16906) doi:10.2196/16906

KEYWORDS

eHealth; blended care; implementation; log data; forensic mental health care

Introduction

Mental health issues cause an increasing number of personal, social, and financial burdens [1] and form a growing challenge for health care systems [2,3]. Technology can be used to address this challenge by supporting treatment of mental health problems in an efficient manner [3,4], while maintaining comparable clinical outcomes as standard in-person treatment [5-7]. The application of technology in mental health care is often referred to as eMental health: the use of technology for treating or preventing mental health disorders [8]. Multiple types of technology can be used. Multimodal web-based interventions based on cognitive behavioral therapies have been studied most often; other examples are mobile apps or virtual reality [8-11]. eMental health technologies can be used as a stand-alone tool, used individually by a person, but often they are integrated within in-person treatment, delivered by one or more therapists. The combination of offline, in-person treatment and online technologies in mental health care is referred to as blended care [12]. Blended care can offer various advantages. Among other things, it has the potential to increase patient engagement and sense of ownership for their treatment, reduce barriers toward receiving mental health care, offer treatment in a more standardized, evidence-based manner, and save time and decrease costs; it can also be personalized to optimally fit patients [4,8,13-15]. However, while eMental health has a broad range of potential benefits, most are not observed in practice [8,16].

An important reason for this gap between the potential and the current situation can be found in issues related to implementation. Implementation of eHealth (electronic health) refers to the strategies that are undertaken to realize the adoption, dissemination, and integration of eHealth innovation into care [17,18]. Examples of such implementation strategies are training and education of stakeholders, changing an organization's infrastructure, using evaluative strategies, or supporting clinicians in using the intervention [19]. Ideally, these implementation strategies have a positive impact on implementation outcomes, defined in Table 1 [20,21]. However, studies show a broad range of issues with implementation outcomes for eMental health interventions, including acceptance by therapists and patients [22], therapists' lack of knowledge on how to optimally combine eMental health and in-person treatment [14], a suboptimal fit with existing technologies such as electronic patient records, and practical barriers such as continuous maintenance of the technology or good internet access [16]. Consequently, to further actualize the benefits that eMental health can offer, implementation strategies should be improved. In order to identify relevant points of improvement, a recent review of eHealth implementation recommended that there is a need for more studies that critically analyze implementation strategies and outcomes of eMental health technologies in practice [23].

Table 1. Implementation outcomes and their definitions, adapted from Proctor et al [21].

Implementation outcome	Definition
Acceptability	Intervention is agreeable, palatable, or satisfactory among implementation stakeholders
Adoption	Intention, initial decision, or action to try or employ an intervention by a care provider or organization
Appropriateness	Perceived fit, relevance, or compatibility of the intervention for a given practice setting, provider, or consumer and/or perceived fit of the innovation to address a particular issue or problem
Cost	Cost impact of an implementation effort, dependent on the costs of the intervention, implementation strategy used, and location of service delivery
Feasibility	Extent to which a new intervention can be successfully used or carried out within a given setting
Fidelity	Degree to which an intervention was implemented as it was prescribed in the original protocol or as it was intended by the program developers
Penetration	Integration of an intervention within a service setting and its subsystems
Sustainability	Extent to which a newly implemented intervention is maintained or institutionalized within a service setting's ongoing, stable operations

Several studies have focused on this issue and identified barriers and facilitators for the use of eMental health in practice [14,24-26]. However, as a recent review pointed out, most of the studies that analyze implementation of eMental health focus on one level (eg, factors related to patients) [16]. In order to get a good grasp of implementation of eMental health, attention needs to be paid to other levels as well (eg, organizational [16] or policy levels [13]). These recommendations on eMental health are in line with more general implementation models and literature: implementation should be seen as a multilevel and

complex process [27] that requires a holistic approach [28,29]. Implementation models like the consolidated framework for implementation research (CFIR) [30] and the nonadoption, abandonment, scale-up, spread, and sustainability (NASSS) framework [31] account for the dynamic interaction between different factors and emphasize the interrelationship between characteristics and perspectives of users, organizations, and the intervention itself. Consequently, analyzing and integrating characteristics and perspectives of users; the context in which the eHealth intervention will be used; and the content, design, and use of the technology itself is expected to result in a complete, realistic picture of the implementation process and outcomes [20,28-32].

In order to apply such a holistic approach to eHealth implementation, a combination of different types of data that provide insight into the different aspects of eHealth implementation is necessary. Collecting multiple types of data does justice to the dynamic, complex interaction between factors that influence implementation, as opposed to analyzing these factors separately [33]. Also, from a holistic point of view, implementation should be studied from multiple angles and perspectives to gain in-depth insight into the technology, context, and people involved [18,28]. To illustrate: if only quantitative data from questionnaires are used to analyze implementation, an in-depth understanding of the reasons for the use of eMental health might be lacking [24]. However, when only using qualitative methods like interviews, information might not be as reliable or objective as is necessary for a thorough analysis of implementation [34]. Consequently, a mixed-methods approach where different types of quantitative and qualitative data are triangulated does justice to the complex integration of factors related to people, technology, and context. This is required for wielding a holistic approach toward the evaluation of eHealth implementation [35-38].

This study applied a mixed-methods approach to the holistic evaluation of the implementation process and outcomes of a blended eMental health intervention introduced in routine care by an organization for forensic mental health care. This setting provides an interesting context to study implementation in practice. First, the evaluation of implementation processes of eMental health technologies that have been implemented in routine care by an organization is expected to result in more ecologically valid results, as opposed to technologies that are being used because of research-initiated studies [20,35]. Second, our study focused on the implementation of an online eMental health platform with multiple modules that has been used for over 4 years in an organization that offers forensic mental health care to both in- and outpatients, which is expected to provide novel insights into long-term implementation processes in practice. Third, forensic mental health care is a branch of mental health care that focuses on treatment of a broad range of in- and outpatients who have committed or were on the verge of committing an aggressive or sexual offense, partly caused by one or more psychiatric disorders [39]. Because of the complex nature of this type of mental health care, and because not much is known about implementation in this type of setting [40], forensic mental health care offers an interesting setting to study implementation strategies and outcomes. Consequently, the goal of our study was to apply a mixed-methods approach to the holistic evaluation of the implementation strategies and outcomes of an eMental health intervention in an organization that offers forensic mental health care. The main research questions are as follows:

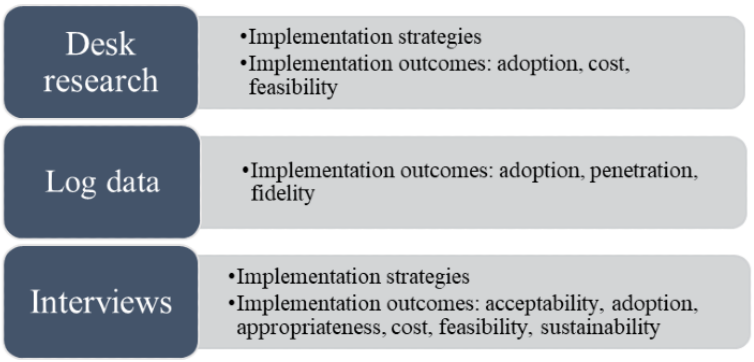
- Which implementation strategies were employed by the organization?
- What are the implementation outcomes in terms of adoption, fidelity, and penetration of the eMental health intervention?
- How do therapists perceive and explain implementation strategies and outcomes in terms of factors related to context, technology, and people?

Methods

Design

This mixed-methods study has evaluated the long-term use of a web-based application from multiple perspectives. A convergent parallel mixed-methods design was used [41] in which qualitative and quantitative data were collected in parallel, analyzed separately, and then merged. First, qualitative desk research was used to describe implementation strategies of the organization. Second, quantitative log data were used to analyze the objective use of an eMental health intervention by therapists and patients to gain insight into implementation outcomes. Third, interviews with therapists were conducted to gain more insight into implementation strategies and outcomes and analyze how they perceive and explain these strategies and outcomes. The purpose of this design is complementary [42]: the qualitative results are used to explain, illustrate, and provide more depth to the results from the quantitative log data, and the quantitative log data are used to enhance and illustrate the qualitative results in order to improve the interpretation of these findings and substantiate conclusions [43]. In order to answer the research questions, results were synthesized in the discussion by means of the aforementioned implementation strategies and outcomes. In Figure 1, an overview of this mixed-method study is provided.

Figure 1. Overview of methods used in this study with types of outcomes seen in each method.



Setting

Organization

This study focused on the implementation strategies and outcomes of implementation of an eMental health intervention within one forensic mental health care organization. This organization started a pilot with the intervention in 2012 and gradually implemented the intervention in the entire organization around the beginning of 2014. The Dutch organization in which this study took place offers forensic mental health care to both in- and outpatients. From January 1, 2014, until May 30, 2019, 3865 in- and outpatients were treated at one of the locations of the forensic hospital. The hospital has two main outpatient clinics, where approximately 85% of patients are treated, and three main inpatient clinics, where the remaining 15% are treated. A total of 252 therapists worked at the hospital between 2014 and 2019.

Electronic patient records show that from January 1, 2014, until May 30, 2019, 2076 patients were treated in the outpatient clinic where the interview study took place, which is 54% of the total patient population of the forensic hospital. According to electronic patient records, 23.27% (483/2076) of the patients had a level of education of primary school or none at all, 22.74% (472/2076) attended secondary school, mostly vocational, 16.33% (339/2076) completed vocational secondary education, 3.32% (69/2076) completed higher secondary education at (applied) universities, and for 34.49% (716/2076), no information was available. Comorbidity was high in this patient population, and there was a broad range of diagnoses for psychiatric disorders (eg, personality, attention deficit, sexual, anxiety, depression, schizophrenia, and substance use disorders).

Online Intervention

The eMental health intervention that is the topic of this study is a website containing a collection of different types of modules. The intervention is suitable for all types of mental health care, not just forensic mental health care. The intervention was designed by a commercial company, and organizations that want to use it must pay for a subscription. In total, 234 modules were available in May 2019. These modules cover a broad range of topics including attention deficit hyperactivity disorder (ADHD), autism, social skills, mindfulness, personality disorders, trauma, addiction, and relaxation. The intervention also contains 6 modules specifically developed for forensic mental health care. These modules focus on creating offense chains and prevention plans, patient recovery, positive self-image, and leading a meaningful life. However, since forensic patients suffer from a broad range of psychiatric disorders and psychosocial problems, other, nonforensic modules are often suitable as well. Therapists must choose which module they find most fitting for their patient; they are able to assign any of the 234 modules. If a therapist does not assign a module, a patient is not able to use the intervention.

Each module consists of multiple sessions provided in a fixed order and accessed via a browser. These sessions consist of a combination of elements (eg, written information about the topic, a story from a peer (in video or text), written assignments derived from cognitive behavioral therapy, and videos to provide additional information about the topic of the session). The underlying assumption is that a patient must complete all sessions in order to be adherent to a module. In our study, the intervention is used as part of blended care, which means that the patient is asked to complete assignments in each session on which the therapist provides written feedback. The patient can only continue with the module once the therapist has provided feedback on a session.

Desk Research

In order to identify the implementation strategies employed by the organization, desk research was conducted. In total, 18 documents describing the pilot project and implementation of the eMental health intervention were obtained from a policy advisor of the forensic organization who has been involved in the implementation of the intervention from the start. Examples of included documents are reports on the planning, progress, and outcomes of the pilot; communication with management; and brief research reports. In order to summarize the implementation process, a timeline with a chronological description of decisions, products, and events was distilled.

Log Data Analysis

Log data from the entire organization from December 2013 until May 2019 were collected and analyzed. These data were analyzed to gain insight into the following implementation outcomes: adoption by therapists and patients, fidelity, and penetration of the eMental health intervention in the organization. Log data refers to anonymous records containing information of every action performed by every user [36]. To be able to analyze the log data, several files with anonymized log data were retrieved from the platform. First, multiple files with information on modules assigned to patients and sessions

completed were downloaded. The raw data were combined and organized into an overview of modules and accompanying lessons by means of a macro in Excel (Microsoft Inc). Second, a file with the monthly number of feedback messages sent by individual therapists was retrieved. All log data were stored and processed anonymously and in line with privacy regulations relevant at that point in time. Ethical approval (No. 18408) was obtained from the ethics committee of the Faculty of Behavioral, Management, and Social Sciences from the University of Twente.

Interview Study

Participants

In order to gain a deeper insight into how therapists perceive and explain implementation strategies and outcomes, interviews were conducted with therapists working at one outpatient clinic of the forensic hospital. In this clinic, therapists were expected to use the eMental health intervention. Therapists were interviewed because of their key role in implementation: if they did not introduce the intervention to the patients, patients could not participate. The attitudes and actions of health care professionals appear to have an essential role in eHealth implementation [44]. At the time of the interviews, 20 therapists were working at the outpatient clinic. All therapists were invited to participate by the manager of the outpatient clinic, but two of them were excluded because they did not receive training and had no experience with the eMental health intervention.

Materials and Procedure

The main goal of the interview study was to identify factors which, according to therapists, are related to the use and nonuse of the eMental health intervention. These factors provide insight into implementation strategies and outcomes. In order to achieve this, semistructured interviews with the 18 therapists were conducted in April and May 2018 by two researchers (KR & NtC) at the outpatient clinic. The interviews were audiorecorded and took between 21 and 61 minutes, with an average of 41 (SD 10) minutes. Ethical approval (No. 18239) for the interview study was given by the ethics committee of the Faculty of Behavioral, Management, and Social Sciences of the University of Twente.

The interview started with a brief explanation of the goal and content of the study. After that, informed consent was signed. The interview scheme consisted of 6 main categories with accompanying open questions. First, sociodemographic questions were asked. Second, experiences with the introduction of the eMental health intervention were discussed. Third, the participant was asked to describe in what way, how often, and with which patients he or she used the eMental health intervention. Reasons for nonadherence were also discussed. The fourth part contained questions on the potential and experienced added value of the eMental health intervention for the therapist, patient, and organization. Fifth, participant was asked to describe what the ideal situation with regard to the use of the eMental health intervention would look like. In the sixth part, barriers for using the intervention were discussed. These questions were divided into 5 topics, loosely based on 5 relevant domains of the NASSS framework [31]: barriers related to patients, therapists, and the forensic health care organization; the wider context; and characteristics of the eMental health intervention. The NASSS framework was used because its holistic nature, in which attention is paid to different types of factors and their interrelationships, fits the research goal of this study. The interview's final question focused on what should be done to overcome these barriers and optimize benefits.

Analysis

The interviews were transcribed verbatim. In order to answer the research questions, deductive, top-down coding via the NASSS framework was combined with an inductive, bottom-up analysis of all fragments belonging to a domain of the NASSS framework. First, all relevant fragments were analyzed deductively by categorizing them into 1 of the 7 domains of the NASSS framework. This deductive analysis ensured a clear main structure of the results in line with the holistic focus of the research goals. The NASSS framework was used to structure the analysis because of its focus on technology in health care and holistic approach [31]. After deductive analysis using domains of the NASSS framework, fragments within each domain were analyzed inductively to look for more specific factors important for the use of the technology according to the interviewed therapists. A coding scheme was iteratively created based on all fragments of the first 5 interviews by one researcher (HK) via the method of constant comparison [45]. Using this coding scheme, the 165 fragments of these first 5 interviews were independently analyzed by a second researcher (FS) to determine interrater reliability of the coding scheme. The joint probability of agreement was 89%. After deliberation of the fragments that were assessed differently by the researchers, agreement was reached on all fragments, and several definitions of the code scheme were fine-tuned. No further adaptations to the underlying structure of the code scheme were required. Because of the high interrater reliability, one researcher (HK) coded the remaining 572 fragments and discussed them with the other researcher (FS) in case of doubt. Again, definitions of codes were adapted throughout the process.

Synthesis

In this mixed-methods study, results were synthesized via the implementation strategies and outcomes. Figure 1 shows which method was used to provide information for which implementation outcome. To synthesize the results, implementation strategies were summarized using all three methods. Also, the most important findings per implementation outcome were described and supported by outcomes of desk research, log data analyses, and codes that arose from the inductive analysis of the interviews.

Results

Desk Research

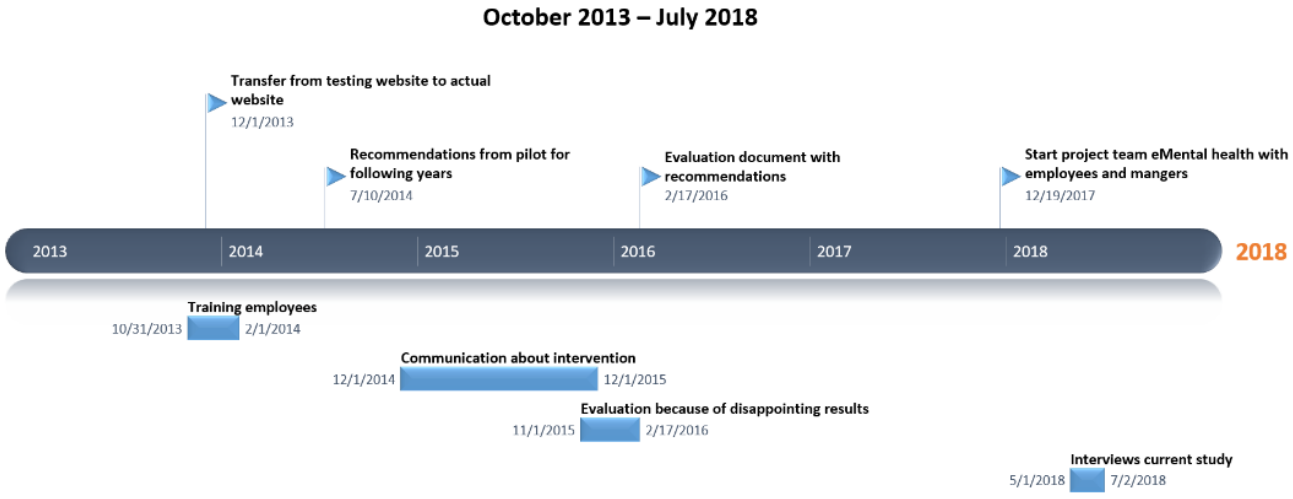
In order to describe the implementation strategies, desk research was conducted with documents generated by the organization. Before the online intervention was disseminated throughout the organization, a pilot was conducted in which the intervention was used on a small scale. This pilot was coordinated by a project team consisting of therapists and policy advisors, and

its timeline is visualized in Figure 2. The goal of the pilot was to improve the content and usability of the intervention and develop a good strategy regarding communication about the eMental health intervention to the organization. The pilot started with an exploratory phase, after which 90 employees were trained and instructed to use the intervention for several months. Desk research did not show how many therapists and patients participated in the pilot. The experiences of the pilot were used to create a strategy and recommendation for implementation of the intervention in the organization. Example of recommendations that arose from the pilot were to provide all employees with the training that participants of the pilot received and write guidelines on how to embed the intervention's modules in existing care programs.

Figure 2. Timeline of the pilot phase of the online intervention.

After the pilot, the online intervention was introduced to all therapists of the organization. An overview of the timeline, until July 2018 when the interview study of this paper was finalized, is provided in Figure 3. The main goal of the implementation was to fully integrate the eMental Health intervention in all primary and supportive processes. Consequently, all therapists received training and were expected to use the intervention. However, because use in practice was not as high as expected, an evaluation was conducted in 2016 that resulted in several recommendations, including that management should improve communication about targets of the eMental health intervention to therapists, a clear overview of useful modules should be created, some skilled therapists should be appointed as champions who can support colleagues in using the intervention, and therapists should motivate patients more by, for example, calling them if they did not use the intervention. When the recommendations did not lead to any major improvements, a new project team to installed to improve implementation in 2017; that team initiated this study.

Figure 3. Timeline of implementation strategies for the online intervention.



Log Data Analysis

Patients

In order to gain insight into implementation outcomes, log data that provide insight into patient use of the intervention were analyzed. From December 2013 until May 2019, 721 unique patients were assigned to at least one module of the eMental health intervention by their therapist. In total, 1019 modules were assigned to these 721 patients. Most patients (514/721, 71.3%) were assigned 1 module, 16.8% (121/721) were assigned 2 modules, 6.4% (46/721) were assigned 3 modules, and 2.5% (18/721) worked on 4 modules. The remaining 2.8% (20/721) worked on 5 to 10 modules. Finally, there were 2 patients who worked on many different modules: one patient worked on 23 modules and another patient on 28. Of the patients, 18.0% (130/721) fully completed at least 1 module, 50.6% (365/721) completed 1 or more lessons but did not complete at least 1 module, and 30.0% (216/721) patients did not complete any lessons at all.

In total, 98 different modules were assigned to patients. The median number of patients assigned to an individual module was 4. The offense script and prevention plan module was assigned to the most patients (104/721), and 18 modules were assigned to only 1 patient. Table 2 provides an overview of all modules that were assigned to at least 10 patients, including an overview of how many patients completed all lessons of the module, completed 1 or more lessons, or did not complete any lessons. When looking at all modules and patients, 180 of the 1019 modules (17.66%) were completed, meaning that all lessons were finished. For 448 of the 1019 modules (43.96%) at least 1 lesson was finished, but not the entire module. On average, when a module was started but not completed, 43.17% (2155/4992) of the modules' lessons were completed. When looking at the longer modules containing 10 to 26 lessons, 44.43% (1994/4488) of the lessons were completed. Of the shorter modules with 9 or fewer lessons, 40.28% (203/504) of the lessons were completed. Finally, in 412 of the 1019 modules (40.43%), no lesson was finished.

Table 2. Overview of the total and relative number of patients that completed, didn't complete, or partially completed modules that were assigned to at least 10 patients.

Topic of module	# lessons	Module completed, n (%)	Module not completed, ≥1 lesson finished, n (%)	Module not completed, no lessons finished, n (%)
Offense script and prevention plan (n=104)	25	14 (13)	71 (68)	19 (18)
Aggression (n=94)	14	7 (7)	48 (51)	39 (41)
Autism psychoeducation (n=76)	10	12 (16)	36 (47)	28 (37)
Substance abuse problems (n=63)	15	5 (8)	37 (59)	21 (33)
Mindfulness (n=59)	9	6 (10)	32 (54)	21 (36)
Offense script and prevention plan (short version; n=55)	17	12 (22)	34 (62)	9 (16)
Expert of yourself (n=53)	10	11 (21)	26 (48)	16 (30)
ADHD[a] (adults): understand your ADHD (n=40)	3	11 (28)	9 (23)	20 (50)
Thought scheme (n=29)	2	9 (31)	10 (34)	10 (34)
Skills for mild intellectual disorders (n=22)	9	3 (14)	11 (50)	8 (36)
Loved ones of patients (n=20)	9	3 (15)	11 (55)	6 (30)
Forensic: positive self-image (n=18)	4	9 (50)	3 (17)	6 (33)
Social skills (n=17)	1	5 (29)	5 (29)	7 (41)
Social skills: saying no (n=15)	2	6 (40)	5 (33)	4 (27)
Information on psychotic disorders (n=14)	9	5 (36)	7 (50)	2 (14)
Psychoeducation for personality disorders (n=14)	4	3 (21)	2 (14)	9 (64)
Generalized anxiety (n=12)	9	2 (17)	7 (58)	3 (25)
ADHD (adults): I want to think before I act (n=11)	9	6 (55)	0 (0)	5 (45)
ADHD (adults): I want to clear my mind more (n=10)	1	2 (20)	0 (0)	8 (80)
Aggression in your relationship (n=10)	14	0 (0)	7 (70)	3 (30)

[a]ADHD: attention deficit hyperactivity disorder.

Therapists

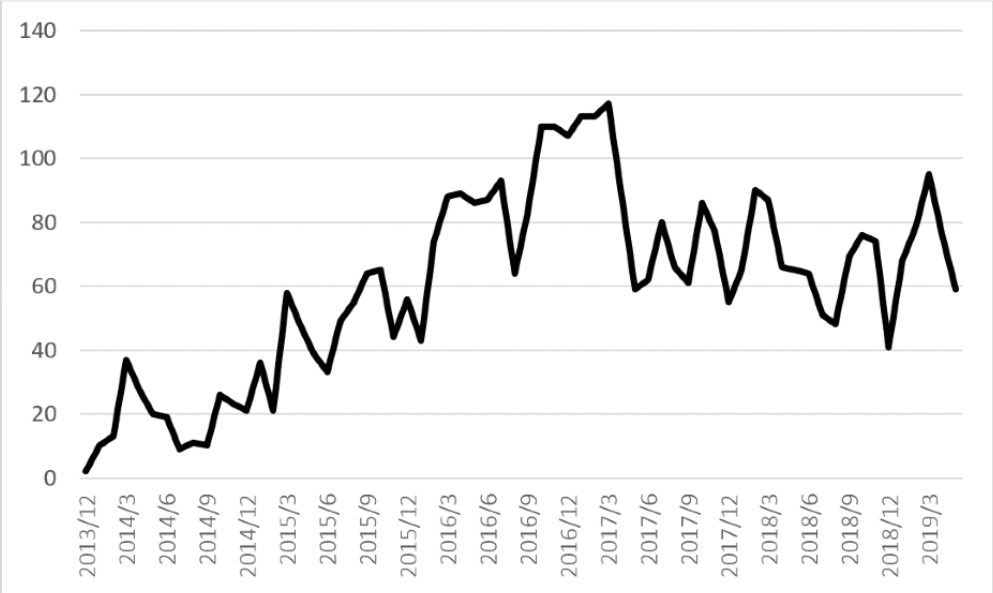
Therapists' use of the intervention was analyzed as well to gain insight into implementation outcomes. A main task of the therapist in using the eMental health intervention was to give feedback on patient assignments. A patient could only continue with the next lesson once the therapist provided feedback, and all lessons required feedback. In total, 172 therapists had accounts, which means they could use the intervention and provide feedback. The median number of feedback messages sent per therapist was 1, with a minimum of 0 and maximum of 460. Of the 54.1% (93/172) of therapists who gave feedback from January 2014 to May 2019, 25.0% (43/172) gave feedback 1 to 5 times, 25.0% (43/172) gave feedback 6 to 19 times, 25.0% (43/172) gave feedback 20 to 50 times, and 25.0% (43/172) gave feedback 51 to 460 times. Table 3 shows how many therapists sent how many feedback messages, showing one major outlier who gave feedback 460 times. The therapist who gave the second highest amount of feedback sent 251 messages.

Table 3. Number of feedback messages sent by therapists.

Messages sent	Number of therapists
0-35	144
35-70	12
70-105	7
105-140	2
140-175	0
175-210	4
210-245	1
245-280	1
280-315	0
315-350	0
350-385	0
385-420	0
420-455	0
455-490	1

Finally, Figure 4 shows the total number of feedback messages that were sent over time, from the introduction of the eMental health intervention until May 2019. The figure shows an increase in sent feedback messages with a peak in 2016, after which the number of messages decreased and seemed to stabilize at approximately 60 messages per month.

Figure 4. Number of feedback messages sent by therapists over time.



Interviews

Participants

In order to gain insight into factors related to implementation outcomes, 18 therapists from one outpatient clinic were interviewed. The therapists had different occupations: 8 psychologists, 6 social workers, 2 system therapists, 1 trauma therapist, and 1 forensic nurse were interviewed. Participants had an average age of 42.5 (SD 10.46) years with a range of 28 to 60 years, and 10 were female. At the time of interviewing, they had been working in forensic care for an average 13.18 (SD 8.68) years with a range from 8 months to 29 years.

First Impressions, Introduction, and Subjective Use of the Intervention

When asked about their first impression of the eMental health intervention and its introduction by the organization, most therapists were positive, as can be seen in Table 4.

Table 4. Therapist responses to survey questions.

Survey question	Very bad	Bad	Neutral	Good	Very good
First impression of the intervention	0	0	6	11	1
Introduction to the organization	0	1	5	10	2

When asked about use of the intervention, 4 therapists indicated that they received training but never used the intervention with a patient. The other 14 therapists had used the intervention with on average 8 patients, with a minimum of 2 and a maximum of 15. Therapists used the intervention in different ways in their in-person sessions with patients. Some discussed the assignments completed by patients very irregularly or never, whereas others discussed them structurally in each session. All therapists saw the intervention as an addition to treatment and used it in a blended manner, where in-person contact was seen as more important. Almost all therapists indicated that most patients did not finish modules and it was hard to keep them motivated.

Coding Schemes

Overview of Codes

The NASSS framework was used to structure the interview scheme and guide the coding process. Consequently, the main codes that were identified in the interviews are mostly aligned with the domains of the NASSS framework, as can be seen in Table 5. However, since a combination of deductive and inductive coding was used, there are several differences between the codes and domains of the NASSS framework. First of all, the adopter domain was split into a therapist and patient code. Integrating all these subcodes into one main adopter code would not have done justice to the differences between these types of factors. Furthermore, the embedding and adaptation over time domain of the NASSS model was not identified in interviews because therapists had difficulties providing a long-term vision on the technology and organization.

Table 5. Main codes, their definitions, number of interviews quote was mentioned (Nint), and number of times code was mentioned (Ntot).

Main code	Definition	N_{int}	N_{tot}
Adopters–therapists	Characteristics, cognitions, or behaviors of therapists that influence the use of technology in treatment	18	203
Adopters–patients	Characteristics, cognitions, or behaviors of patients that, according to therapists, influence the use of technology in treatment	18	194
Value proposition	The (desired) added value that a technology has or should have for treatment, according to therapists	18	121
Technology	Influence of the technology's features on use by therapists and patients	18	104
Organization	Characteristics, culture, and activities of the organization that influence use of technology by therapists	18	94
Wider system	Influence of activities by the broader context on use of technology by organizations and therapists	11	7
Condition	Extent to which the nature of the patient's condition (or disorder) influences use of technology, according to therapists	3	4

Adopters–Therapists

This main code was mentioned most often and refers to characteristics, cognitions, or behaviors of therapists that influence the use of eHealth in treatment. This code was found in all interviews, and 203 fragments belonging to this main code were identified. The subcodes are reported, defined, and illustrated with one or two quotes in Table 6. As can be seen in the table, therapists discussed a broad range of factors related to themselves that could influence the intervention's use. The most mentioned code referred to them not putting in enough time or effort to start or keep using the intervention. This could result in a lack of knowledge and skills to optimally use the intervention. The perceived lack of time was partly attributed to high workload but also to lack of enthusiasm to use the intervention. Among other things, not all participants were keen to work with technology in general. Several therapists believed that the intervention did not have enough benefits for their patients, and some felt that the technology was not easy to embed within standard treatment. Related to this, many therapists indicated that using the intervention was not in their system: while they often were willing to try it, most of the time they simply forgot about it because it was not part of their treatment routine. Consequently, many therapists did not introduce the intervention to their patients or did not motivate them enough to keep using it. Because the intervention was often not on the top of their minds, therapists indicated that they hardly discussed it with their colleagues. Finally, according to several participants, an important reason for successful use was having experienced benefits of the intervention for the patient: if the intervention fits an individual patient's skills and problems, chances on successful use were said to be higher.

Table 6. Subcodes of the main code "adopters–therapists," their definitions, illustrative quote, number of interviews the quote was mentioned (N_{int}), and number of times code was mentioned (N_{tot}).

Subcode	Definition	Illustrative quote	N_{int}	N_{tot}
Investing effort and time	Effort and time the therapist is able and/or willing to invest in getting acquainted with and structurally using the technology	• *You have to know what happens in the module, what it contains. And then you have to check which assignments have been completed, and then give feedback on it. That costs extra time. Perhaps that's why people are hesitant. (pp. 13)* • *It can't be like you saying: "Hey, just fill it in, but I'll never look at it again." You have to make time to thoroughly look into it and to familiarize yourself with it. (pp. 4)*	18	48
Supporting patients	Extent to which a therapist actively introduces the technology to patients and/or tries to motivate and support patients to keep using it	• *Sometimes I know in advance: "This one will not use it," and maybe that's a self-fulfilling prophecy, that might be. But still, I'm not really inclined to start then. (pp. 7)* • *Because I've tried a lot of times to start with a certain module with patients who are not passionate about it. [...] But I notice it with myself: because the passion is there I can often make it work. So that's very necessary. (pp. 12)*	17	37
Integration of technology in routines	Extent to which the technology is integrated in the therapist's routine and/or whether they automatically think of using the technology	• *Yeah, I don't know what's the reason, I cannot really explain it, but despite that I like the modules, it's not on the top of my mind. So there is something that's stopping me from diving into my computer to find out which module fits with a patient. And I don't really know why it doesn't come up. (pp. 15)* • *Well, I agreed with a patient that they would hand it in next week Tuesday. Well ok, but then I didn't receive it and I forgot about it myself. The patient stopped doing it and well, it kind of disappears into the background. I now resolved it by planning appointments in my calendar, so that I'll be reminded to send reminders to patients. (pp. 6)*	16	36
Knowledge and skills	Therapist's level of knowledge about the technology and skills to appropriately use it in treatment	• *I also find it difficult to figure out which module fits which patient. [...] And I think that in my case, I'd just have to know which modules exist. And then I can just say "This fits you nicely, I'd like to recommend this, see if it's possible" and then get to work. (pp. 7)*	13	34
Attitude toward the technology	Therapist's opinion on and feelings toward using technology in treatment	• *I think it depends on in which generation you've grown up, I think younger colleagues have more feeling with it. [...] and that asks for a considerable adjustment, transformation, also for a number of older colleagues. (pp. 1)* • *And working with people, I still like doing that. Except for all that technical fuss, computers mean nothing to me, and also smartphones... I don't have any feeling with that and experience it as a burden. (pp. 2)*	12	28
Discussing technology with colleagues	Technology as a topic of conversation among therapists inside and outside of official meetings	• *Then I'll say in our team: "I'm hearing this and this and this," and sometimes someone in the team says: "Hey, eMental Health is something." I think that if people in our team say "Maybe eMental health is an option" more often... And if that rhythm isn't there, that's the reason why there is very little eMental health. Or at least within our team. (pp. 12)*	10	21
Experienced benefits	Extent to which a therapist perceives that the technology has positive effects on the patient's treatment outcomes	• *So then we try to shape the treatment in another way, instead of continuing with something [the module] of which you have the idea that it doesn't have that much effect. (pp. 3)*	2	4

Adopters–Patients

The main code "adopters–patients" was mentioned by all 18 therapists and refers to characteristics, cognitions, or behaviors of patients that influence the use of eHealth in their treatment. In total, 8 subcodes, presented in Table 7, and 194 fragments related to this main code were found. Therapists discussed multiple types of patient-related factors that, according to them, could influence the use of the intervention. It was frequently mentioned that many forensic psychiatric patients are often not motivated to start or keep working on the intervention. Therapists indicated that this was in line with low motivation for their treatment in general, which is partly due to the often obligatory nature of these patients' treatment. Furthermore, therapists stated that a large share of the patient population has cognitive impairments due to psychiatric disorders—such as problems with focusing—or received very little education (eg, having finished only primary school). According to the therapists, this can cause problems with patients understanding the mostly text-based intervention, completing written assignments, being able to individually reflect on their behavior, and having the required technological skills to be able to practically use the intervention. Furthermore, patients have to work in the intervention individually in their own time, which some therapists compared with homework. Therapists stated that many patients have difficulty with this: a large share of the patient population was not seen as conscientious enough to independently work on the intervention. For example, patients

often do not stick to agreements about when assignments were to be completed. Furthermore, therapists indicated that forensic psychiatric patients often have multiple psychiatric disorders and problems within their social environment, which might negatively influence their use of the intervention. When severe psychosocial issues occur, patients might be too preoccupied with these issues to use the intervention. Furthermore, not all patients have access to a computer or laptop, or they do not have a quiet place where they can comfortably work on the intervention.

Table 7. Subcodes of the main code "adopters—patients," their definitions, illustrative quote, number of interviews quote was mentioned (N_{int}), and number of times code was mentioned (N_{tot}).

Subcode	Definition	Illustrative quote	N_{int}	N_{tot}
Motivation	Extent to which patient is motivated, enthusiastic, or open toward working with the technology in their treatment	• *I kind of think that, I think I've mentioned before that if that motivation is present, if they have the feeling that eMental Health fits their problems. And that has to be the case. They don't have to think: "Well, what about the problems I have...." They do have to see the connection to be motivated. (pp. 4)* • *Because very demotivated patients who don't want to be here anyhow... See, it's easier to not work on eMental Health at home, then not attending a face-to-face appointment. They'd earlier drop the eMental health then not coming here. (pp. 9)*	14	40
Conscientiousness	Extent to which patient is diligent in working on the technology and fulfills commitments regarding the use of the technology outside of treatment	• *Reading comprehension, making assignments, those kinds of things. If you're not used to doing homework, if you've never done homework in your life—eMental Health is actually homework. So that's a skill in itself. They have to be able to do it. (pp. 9)* • *Next, we open the module and then it seems to go well for about 1 or 2 sessions. But then there's always an "appointment forgotten" or "not finished" or those kind of things. After 3 months someone has stopped filling in the assignments, and only completed two sessions. (pp. 11)*	14	27
Literacy and educational level	Patient's ability to write, read, and understand treatment-related information in the technology	• *But I noticed that it wasn't really working. We've done it a couple of times, but there were a lot of difficult words. And if I explain the words, he'll forget what it means after that. (pp. 4)*	14	22
Experienced benefits	The extent to which a patient experiences a positive influence on his or her treatment because of the use of the technology	• *And whether the module fits the treatment and problems of the patient. [...] If you, for example, have a module on autism, and someone recognizes himself in that, in a module on psychoeducation for example. Then it can actually be useful in their daily lives. (pp. 6)* • *And also what I said earlier, that a module fits the needs of a patient. So if you use a sleeping module on someone with sleeping problems, chances are higher that he will continue with it. (pp. 16)*	14	22
Psychosocial situation	Level of stability of patient's personal life and/or mental state that is required to use a technology	• *I don't really think about it with people who are in a crisis situation. Because in my experience, they really don't have the need to discuss [the intervention]. You have to be in clearer waters, before they at least... Yes, that's my experience, before you can make a connection with someone who's overwhelmed by stress. If you, for example, have trouble about your social security benefit, or trouble with the neighbors, or trouble with your spouse, then it doesn't really work for those people anymore. (pp. 10)* • *Patients who are very prone to psychiatric crisis [...] or have so many psychosocial problems, well, then it won't work. Then you're only trying to stabilize them and you cannot use eMental health. (pp. 9)*	13	28
Technological skills	Level of practical skills required for successfully using information and communication technologies such as computers or smartphones	• *And I also notice that with eMental Health, the problem is that if a patient doesn't know anything about the computer, who isn't focused on that, you won't be able to make it work. You can jump high or low, but you won't get it done. (pp. 12)*	12	17
Availability of technological resources	Patient's access to necessary preconditions to use the intervention: technological device, appropriate working area, and good internet connection	• *I think that in their own environment, where they like doing it. They have to be able to do it privately, not that there's someone around the entire time. So privacy is important for them, I think. We can't facilitate that; they have to arrange that themselves. Or we'd have to offer them a place to work here, so they can sit behind a computer here. (pp. 7)* • *But I have had several people who were pretty positive about it, but who didn't have a computer, or their computer broke. (pp. 8)*	11	18

Subcode	Definition	Illustrative quote	N_{int}	N_{tot}
Reflective skills	Patient's ability to independently write about and reflect on emotions, cognitions, and behaviors in the technology	• *I'm dealing with a lot of patients that find it really difficult to put their emotions and feelings on paper. They don't talk about those things regularly, like conflicts with their wife. Those people just didn't learn that. And yeah, if that's already difficult in a conversation to tell what you feel, what you want, or what you want differently, it's even more difficult to type it if you're alone, sitting behind a computer. That's on another level. (pp. 10)* • *Because the simplest question—he understands them, but he said to me: "I simply cannot put it into words." And if you have to do such assignments, and answer things... The fact that he hardly gives an answer, that partly has to do with him not being able to visualize and verbalize. (pp. 1)*	11	21

Value Proposition

The main code "value proposition" refers to the added value that a technology has or should have for treatment, according to the therapists. It was mentioned by all 18 therapists, 5 subcodes were identified, and 121 fragments were found in all interviews. As can be seen in the previous tables, not all therapists were positive about the intervention and did not use it often, but they were able to identify a broad range of potential and actual advantages of the eMental health intervention. These focused, among other things, on the content of the treatment: the intervention was said to have the potential to improve the quality of treatment by, for example, providing more structure to the treatment. Also, because therapists often also read the text of the intervention, several participants indicated that using the intervention might further improve or deepen existing knowledge about disorders. The intervention can also support patients in gaining new knowledge and skills (eg, new insights about a psychiatric disorder or an improvement of reflective or coping skills). Furthermore, several therapists explained that because patients must work on the intervention individually, their feeling of responsibility for their own treatment might increase, and they might ascribe positive changes more to themselves instead of their therapists. Moreover, several practical advantages were mentioned, among which saving time of therapists and patients because of less traveling time and replacing part of in-person treatment with the intervention, an increase of patients' access to care because they can individually work on their treatment at their own pace, and providing a new way of delivering treatment to patients. The subcodes are presented in Table 8.

Table 8. Subcodes of the main code "value proposition," their definitions, illustrative quote, number of interviews quote was mentioned (N_{int}), and number of times code was mentioned (N_{tot}).

Subcode	Definition	Illustrative quote	N_{int}	N_{tot}
Improving treatment	Possibility of technology to improve quality of and further structure of face-to-face treatment	• *It does help you to focus the treatment on what someone needs. And it helps to not get bogged down in other, less relevant matters that people bring up, where it actually shouldn't be about. Someone has a treatment goal and the program fits that goal and that is what you will be doing. So it really helps to frame your therapy. (pp. 11)* • *Often, there are also things in there that I don't have on the top of my mind. And sometimes it's pretty nice to work via a protocol, that you encounter things of which you think: "Hey, I didn't think about that at all!" (pp. 2)*	14	39
Practical benefits	Benefits for patients, therapists, and the organization related to practical matters such as time and money	• *Well, I think it's very nice that you can put patients to work at a time which suits you, and that they can work whenever it suits them. So planning appointments is less of a hassle. And it's just like communication via WhatsApp or email: you all do it in your own time, so in that sense it's easier insertable in everyone's schedule. (pp. 16)* • *Well, the patient can work on it at home. He'd have to come here less often, face-to-face. So you'd have to plan an appointment less often, which can enable you to see more other patients. (pp. 5)*	14	26
Increase of knowledge and skills	Possibility for the patient or therapist to acquire new insights and skills into the patient or the disorder	• *An advantage for a patient with whom I have done the aggression module is that he did really gain more self-insight and came a little closer to himself. With the other patients I didn't really have a hallelujah experience, but with him it seemed like the penny has dropped. (pp. 15)* • *What I like about that is that you also go through the content yourself. [...] And I really like that of the modules, that you learn things from it yourself, and him as well. (pp. 12)*	13	22
Increase in patient independence	Enabling patients to work more independently and feel more ownership for their treatment	• *Well, the fact that someone does it himself with the module, that they are prouder at themselves because they've achieved something. That it wasn't the therapist who helped you, but that you've done it yourself. (pp. 16)* • *If they're in my room, I often have a fairly high pace, I am pulling and pushing them. But with the intervention, they're in a calm environment and can think calmly. (pp. 2)*	13	19
More options for treatment	Possibility of technology to offer a broader range and different types of treatment to patients	• *Well, for example, if a patient is working on the offense chain, you might notice that he has the need to practice more skills. Or wants to read a bit more about certain relaxation exercises. And if we're not there yet in the module, I move these topics forward, so that we can work on those at that point in time. Or I check in another module if there's something there, that they can work on a topic in between. [...] Then I think: "This fits well at this point in time," and then we can continue with the treatment. (pp. 9)*	11	17

Technology

This main code was mentioned in all 18 interviews and focuses on the influences of the technology's features on use by therapists and patients. In total, 3 subcodes and 104 fragments were identified, as can be seen in Table 9. Usability of the technology was often mentioned by therapists. While a few were fairly positive, most found the technology not easy to use for themselves or for patients: it did not fit their preferences and way of working. Furthermore, while most therapists were relatively positive about the intervention's look and feel, it was mentioned that the way the content was presented was not very suitable for many patients, for example, due to a lot of text or too many sessions within modules.

Table 9. Subcodes of the main code "technology," their definitions, illustrative quote, number of interviews quote was mentioned (N_{int}), and number of times code was mentioned (N_{tot}).

Subcode	Definition	Illustrative quote	N_{int}	N_{tot}
Ease of use	Extent to which therapists find use of the technology intuitive, clear, and structured	• *I feel that the eMental health intervention is too big, or too fuzzy. As I just mentioned, all these modules, I know a few, but I think that there are a lot, also specified on other diagnoses. But I find it tricky to find these things. So I don't think it's very well arranged. (pp. 9)* • *I'd see that as tiles in [the electronic patient record] which we are using currently. In User you have multiple tiles and the patient record, and it should also have a tile of [the intervention], on which you click and then you can start. [...] If there would be a block of [the intervention], it would be really easy to go to it. I think that would be more user friendly than when you have to go to the website yourself to log in, because then you have taken multiple additional steps. (pp. 7)*	16	68
Presentation of content	Therapist's opinion on the ways in which the treatment-related content of the technology is presented to the patients	• *The module itself should be shorter. Both the individual sessions and the number of sessions in a specific module I'd make shorter. I feel that certain explanations are too difficult for some patients. So I think there is not enough supply for people with a low intelligence, and a large number of our people has a lower than average intelligence. (pp. 18)*	8	21
Appearance	Therapist's opinion on the overall look and feel of the design of the technology	• *Because I think those modules are really cool. I'm thinking "Wow, the person that came up with this has it right!" But it's just fresh, I'd almost say happy, but also friendly, and user-friendly. And if you can have these things together in a module text, with some videos and some other things and some explanation, that's just amazing! (pp. 12)* • *With regard to the design, with videos and images, it's stimulating and appealing. It's not a boring booklet that you hand out. (pp. 4)*	7	10

Organization

This main code refers to the characteristics, culture, and activities of the organization that influence the use of technology by therapists, and was mentioned in all 18 interviews. In total, 94 fragments for 4 subcodes were identified, which are explained in Table 10. Almost all interviewed therapists explained that the intervention was introduced to them by means of a course in which they gained practical skills to use it. According to therapists, the organization did not pay a lot of attention to the intervention after this course. Multiple therapists indicated that the intervention was often not discussed in official meetings. This was viewed as a partial explanation for therapists not remembering to use the intervention on a regular basis. Also, several therapists indicated that they did not experience enough support for questions about the content of the intervention or the way they could embed it in treatment. To illustrate, some therapists required more support when working with unmotivated patients or had questions about how to integrate assignments of the intervention in their in-person treatment sessions. Besides content-related support, therapists also mentioned several practical barriers that the organization should address, such as a slow internet connection.

Table 10. Subcodes of the main code "organization," their definitions, illustrative quote, number of interviews quote was mentioned (N_{int}), and number of times code was mentioned (N_{tot}).

Subcode	Definition	Illustrative quote	N_{int}	N_{tot}
Introduction of technology to therapists	Activities that the organization undertook to introduce the technology and train therapists' necessary skills and knowledge, according to therapists	• *Hmm, much less, because, I think I had a course once, just in the beginning, and there's not really a follow-up. (pp. 10)* • *We've received a very clear explanation, I believe an entire day and you just started practicing. And I liked that. But now I'm thinking, if I'd have to use it with a patient, I'd really have to ask with the people who gave the course. He'd have to quickly explain to me how it works. That knowledge has faded. (pp. 1)*	17	27
Providing support for therapists	Ways in which the organization offers content-related support to therapists for using the technology in treatment	• *I just think that, well, for the long term the organization has to pay more attention to it, for the team. And I also think that you have to implement the trainer-trainer idea. That you pick a couple of people who work well with it or are a bit better in it, that they are appointed as a source of information, and that other people know about that. (pp. 9)* • *Currently I'm trying to figure it out myself, but how do you really shape a blended treatment? If there would be education about that! Also about the more challenging cases, if there is very little motivation. (pp. 4)*	14	34
Integration in organizational structures	Extent to which a technology is structurally featured in activities or products for which the organization is responsible, such as meetings, treatment protocols, targets, or performance reviews	• *But it doesn't get indicated a lot in the intake or in meetings. Often the one who's done the intake has to bring it up themselves, but it's not something that other people in the meeting bring up or come up with, that that's also an option. My personal experience. If I bring it up, they say "Oh yeah it's a good idea," but if I say nothing, they hardly ever come up with it. (pp. 18)* • *So that management says: "eMental health guys, don't forget about that!" (pp. 6)*	11	21
Providing necessary conditions for use	Extent to which the organization ensures boundary conditions such as availability of sufficient technological resources and time for therapists	• *Internet is often slow. Then you've planned half an hour, and you think, I'm going to do it. Well, it doesn't work. It's also a technical issue. From the division around it, the internet, that's so slow. And if you have to start a program such as [the intervention], well, you can do something else in that time... Then you constantly have to wait, well, then I drop out. (pp. 2)* • *And that time is actually scheduled for people. You have to really work on it, so it cannot disappear in the other activities, because you will forget it. If you really have an hour to only work on eMental health, it will remove some barriers and other arguments. (pp. 18)*	7	10

Wider System

This main code refers to the influence of activities by the broader context on use of technology by organizations and therapists and consists of 2 subcodes, which can be found in Table 11. This code was identified 17 times in 11 interviews. Therapists discussed the wider system less often than previous codes. Several participants briefly mentioned health insurance companies and government but did not elaborate on the role of the wider system.

Table 11. Subcodes of the main code "wider system," their definitions, illustrative quote, number of interviews quote was mentioned (N_{int}), and number of times code was mentioned (N_{tot}).

Subcode	Definition	Illustrative quote	N_{int}	N_{tot}
Demands of health insurance companies	Therapist perception about financial incentives for using the technology offered by health insurance companies	• *I don't know, that's something from the health insurer, that they say that we have to do something with [the intervention]. That's a bit how it feels. That we received that assignment because we have to meet the numbers. (pp. 7)*	9	12
Encouragement of government	Extent to which use of a technology is encouraged by the government	• *First, they should make a statement as an organization, together with other organizations, to say: "Yeah it's all good and you can want it in this way but we're going higher to the government and say to the government: this is not ok." There are so many administrative tasks that don't... If you do something and you have to account for this it's fine, but it can be a lot easier. (pp. 12)*	4	4

Condition

This code refers to the extent to which the nature of the patient's condition (or psychiatric disorder) influences use of the technology according to therapists. The subcode was mentioned 4 times in 3 interviews, as can be seen in Table 12. Therapists often did not discuss their patients' diagnoses as a separate factor that directly influences the intervention's use.

Table 12. Subcodes of the main code "condition," their definitions, illustrative quote, number of interviews quote was mentioned (N_{int}), and number of times code was mentioned (N_{tot}).

Subcode	Definition	Illustrative quote	N_{int}	N_{tot}
ADHD[a]	Impact of (symptoms of) ADHD on patient's use of the technology	• *So I cannot really... I do know that those ADHD patients, that it takes too long for them. They're too easily distracted, or find a question too difficult. It's not that everyone has that. (pp 2)*	3	4

[a]ADHD: attention deficit hyperactivity disorder.

Synthesis

When looking at the implementation strategies described by Waltz et al [19], desk research and interviews showed that in this study, most attention was paid to training of therapists, as mostly becomes clear in the subcode *introduction of the technologyto therapists* of the main code "organization." However, little to no strategies related to changes in the organizations' infrastructure, engagement of patients, and adaptation of technology to the context were conducted. This is further illustrated by all subcodes of the main code "organization," but also by the therapist- and patient-related subcodes *integration of technology in routines*, *knowledge and skills*, *discussing technology with colleagues*, and *motivation*. Furthermore, desk research showed that support and assistance for therapists was available, but most interviewed therapists did not experience this as such, which becomes most clear in the "organization"-related subcode *providing support for therapists*. Desk research showed that besides training, several relatively minor evaluation strategies were conducted by the organization itself, during and shortly after the pilot. However, the outcomes of these evaluations did not lead to major changes to the implementation strategies, so no lasting improvements in the use of the intervention were observed in the log data. This is visualized in Figure 4; a short peak in sent messages can be observed during the time of the evaluation, but this increase in sent messages only lasted several months. To conclude, implementation strategies were mostly focused on training of therapists, but little attention was paid to adaptiveness of the technology, changes in the organization, and patient awareness.

The results of the desk research, log data analysis, and interviews were used to assess the implementation outcomes described by Proctor et al [21] from a holistic perspective, structured via the NASSS framework [31]. First, the interviews showed that acceptability of the intervention was relatively high, with therapists being positive about the intervention and able to identify its added value. This fairly high acceptability is illustrated in Table 4, which shows overall good first impressions of the intervention, and is further supported by the main code "values," which points out that therapists are able to mention a broad range of potential and actual advantages. However, despite the fairly positive acceptability, log data analyses and interviews clearly showed that adoption was low; a large share of therapists and patients did not use the intervention at all. Furthermore, the intervention's penetration in the organization was low; log data showed that only a small fraction of therapists and patients used the intervention. Only 54% of the eligible therapists actually used the intervention, and Figure 4 shows that the largest share those who did use it, did not use it a lot. When the intervention was used, fidelity was often low, as can be seen in Table 2. Only 18% of the modules were fully completed, and of the remaining 82%, modules were either not started or not fully completed, implying that they were not used as intended. An explanation for this can be found in the appropriateness of the intervention. Log data showed that several modules were completed, and several therapists indicated that they were able to successfully use the module with some patients. However, therapists indicated that the intervention did not optimally fit most patients' skills and preferences. This is illustrated by the patient-related subcodes *conscientiousness*, *literacy and education level*, *technological skills*, and *reflective skills*. The mismatch between patient characteristics and the intervention also becomes clear in the main code "technology," which shows that usability, design, and content are not optimally tailored to the forensic psychiatric patient population. Furthermore, the intervention also seems not to be appropriate for many therapists, who indicated that they prefer in-person contact and often felt not fully equipped to integrate the intervention in their treatment. This means that currently, the intervention's costs in terms of finances and time investment seem to be higher than the benefits. Also, sustainability was low: therapists stated that the intervention was often not discussed in meetings and was not integrated in electronic patient records they used. While this theme appears in multiple codes, it becomes especially clear in the organizational subcode *integration in organizational structures*. Consequently, therapists often did not even think of the possibility to use the intervention in treatment, which is represented by the therapist-related subcode *integration of technology in routines*. Currently, the feasibility of the intervention is low because of a suboptimal fit between the features of the technology; needs, wishes, and skills of therapists and patients; and characteristics and activities of the organization. It appears that since the implementation strategies were not conducted from a holistic perspective but mainly focused on training the therapists, the implementation outcomes are disappointing.

Discussion

Principal Findings

Main Outcomes

This mixed-method study evaluated the implementation strategies and outcomes of an eMental health intervention in forensic mental health care from a holistic perspective, where attention is paid to factors related to people, organizational context, and technology. Triangulation of the outcomes of desk research, log data analyses, and interviews with therapists

showed that the technology did not optimally fit the therapists, patients, and organization. Furthermore, the implementation process was mostly focused on skill training of therapists and not executed from a holistic perspective; not enough attention was paid to changes in the organization, patients, and other required changes in therapists. The results of this mixed-methods study will be discussed in more detail structured by the main elements of the holistic approach that was applied: the people using the eMental health intervention, the organization in which it was used, and the technology.

Therapists

The interviews and log data showed that although several therapists were active users of the intervention, most of them only tried it once or twice, and a relatively large share of the therapists did not even use the intervention at all. Nevertheless, almost all interviewed therapists were fairly positive about the intervention and able to identify its added value. This shows that cognitions, intentions, and feelings of users are not fully predictive of successful use. Nevertheless, models that focus on individual factors predicting technology acceptance, such as the technology acceptance model (TAM) [46] or the unified theory of acceptance and use of technology [47], are still used regularly to analyze or plan implementation. While these types of models are useful to create an overview of individual beliefs and attitudes that influence a person's intention to use a technology [48], they pay little to no attention to influences of the context, characteristics of the technology, and interrelationships between them [48,49]. Consequently, implementation models or frameworks that apply such a holistic approach like CFIR [30] or the NASSS framework [31] seem to be more useful in this context because of their focus on a broad range of contextual and (inter)personal factors and not merely individual factors of end users.

Despite the fact that all therapists in the organization had received training, only a relatively small proportion actively used the intervention. This implies that skills training only did not suffice for successful implementation: more than just a how-to instruction seems to be necessary to fully equip therapists to embed the intervention in their treatment sessions. Among other things, therapists also need to know how to persuade patients to start with the intervention, they need to be able to keep motivating patients to complete exercises, and they must embed the content of the intervention and the patient's answers in treatment [14,50,51]. This implicates that the use of eMental health might also change the role of the (forensic) mental health professional [51]. In this new way of working, patients might be more in the lead and supported by professionals, and the structure and content of treatment may not be determined only by the professional but also by the intervention. Such a technology-induced role change in domains where communication previously only took place between persons requires changes on a multitude of levels, like management, education, or government [52], and not merely a skills training of therapists. More research on the nature of this role change and implications for implementation strategies is required.

When looking at implementation strategies, it might also be useful to conduct more research on the need to better tailor these strategies to different types of therapists. In this study, there appeared to be a lot of differences in therapists regarding their subjective attitudes and objective use of the intervention. This is in line with a recent study that showed that therapists differ in the types of drivers and barriers they perceive with regard to the use of eMental health [24]. This might imply that different types of therapists benefit from different types of implementation strategies. For example, therapists with a low level of enthusiasm and skills might need to receive a different type of training than enthusiastic and tech-savvy therapists. In line with this, multiple researchers stated that one-size-fits-all interventions are not very suitable for forensic mental health care and that tailoring is advised [40,53-57]. This argument can be extended to professionals: adaptive implementation strategies that fit different types or subgroups of therapists' needs, skills, and attitudes might be beneficial for implementation outcomes. Subsequent research might focus on the identification of different subgroups of therapists, for example, in terms of attitude or eHealth literacy and tailoring implementation strategies to these characteristics. Also, researchers should assess whether tailoring implementation strategies to different types of professionals actually results in better implementation outcomes.

Patients

Therapists indicated that the eMental health intervention requires a relatively high level of reading and writing skills, cognitive reflection, and conscientiousness. However, according to desk research and other literature, forensic psychiatric patients often have low education levels [58,59], which might explain the low number of patients that completed modules. This shows that there seems to be a poor fit between most users' skills and the content of the eHealth technology, which might be a major cause for nonuse or nonadherence [60,61]. Several solutions might address this issue: therapists can support patients more in working on difficult elements of the intervention, or texts and assignments can be shortened and made easier. However, it might also be possible that the studied eMental health intervention is not very suitable for this context and another type of technology would be a better fit for most patients. For example, multiple recent studies point out the potential of interactive virtual reality interventions for forensic mental health care [62-66], among other things because they allows patients to actually practice with behavior instead of talking or writing about it. Another possibility is wearables, which can be used to collect physiological data associated with aggressive outbursts or as electronic momentary assessment devices to gather information about a patient's emotional state [67,68]. This study underlines the importance of adaptability of eMental health to optimally fit the needs and characteristics of individual patients.

Organization

As was mentioned before, the organization focused implementation strategies mostly on skills training of therapists but did not pay much attention to the implementation of the intervention on other levels. The disappointing implementation outcomes show the importance of the use of multiple types of implementation strategies to ensure that an eMental Health intervention is thoroughly embedded in a forensic organization's

infrastructure [40,50]. This is in line with literature on eHealth implementation in general, which emphasizes the importance of integrating technologies in existing organizational structures or even changing the way care is delivered or organized [28,29,32,33,69]. One way to achieve this in the studied organization is by ensuring that therapists structurally discuss the possibility of using an eMental health intervention at predetermined moments in treatment (eg, during a patient's intake). This can be done by means of the existing "fit for blended care" instrument [12], which aims to support therapists in shaping their blended treatment in cooperation with the patient. This instrument can be adapted to fit the specific forensic mental health organization by means of the patient-related factors identified in this study. Furthermore, the eMental health intervention might need to become a permanent item on the agenda of team meetings to ensure that it is discussed regularly [50]. Moreover, therapists indicated that they hardly discussed the intervention with colleagues, as opposed to other parts of their treatment, so peer-coaching sessions to discuss the use of the intervention might be organized [50]. It is important that research is conducted to determine whether these types of strategies actually boost implementation outcomes [23]. Generating more knowledge on suitable and successful implementation strategies will support other organizations in planning implementation and prevent them from reinventing the wheel, which will eventually save time and money.

Technology

One reason for the low use of the intervention was that not all therapists were positive about the user-friendliness of the intervention's design: among other things, they indicated that the website did not give them a clear overview of suitable modules for specific patients. Log data indeed showed that only a fraction of modules were used frequently. Adding more persuasive elements to the intervention might support therapists in using the intervention. An example is tunneling: the system can guide the therapist through the process of selecting suitable modules for a patient [70]. Additional research can be conducted to evaluate and improve the persuasiveness of the intervention (eg, by means of the Perceived Persuasiveness Questionnaire [71]), which might increase its use [72].

A characteristic of the intervention that might have hindered use is the lack of possibilities for personalization. As was mentioned before, the technology does not seem to fit most patients according to therapists: there were too many sessions within a module, there was too much text, and the subject matter was too complex for most patients. Therapists expressed the need for multiple versions of the modules in order to personalize the intervention. Examples are the possibility to choose between videos or text or the option to select texts with different levels of difficulty. Studies on eHealth in general have stated that personalization can increase adherence [73-75], so a more personalized version of the intervention might result in better implementation outcomes. However, more research is necessary on what elements should be personalized, how this should be done, and if this actually positively impacts use and adherence. Ideally, this redesign of the technology should be done in close cooperation with end users to ensure that it better fits their needs [29,64], since cocreation can also have a positive influence on implementation outcomes [28,76].

Strengths and Limitations

This study took place at one forensic psychiatric hospital in the Netherlands, which might raise questions about the generalizability of the results. However, other studies on eHealth in forensic care have identified similar types of implementation issues [50,53,56,63,77]. On top of that, many of the identified issues have been reported for eHealth in general (eg, lack of enthusiasm in therapists, low adherence by patients, or lack of integration of a technology in an organization's structure [32,33,69]). This implies that, on an abstract level, our findings are relevant for other types of (mental) health care as well. A related strength of this study is that all therapists working at the outpatient clinic could be interviewed, which prevented a self-selection bias from occurring.

Furthermore, while use of the NASSS framework to structure the qualitative analysis was a strength of this study, several issues arose during the coding process. A chief example of this is the domain "condition." As opposed to a patient population that suffers from one disorder (eg, depression or diabetes), the forensic psychiatric population is not characterized by one condition: comorbidity is very common among forensic patients, and patients have committed a broad range of offenses [78-80]. This raises the question on how to characterize certain behaviors or cognitions: as part of a patient's personality or as a symptom of a disorder. To illustrate: if a patient has trouble focusing, which hinders the use of an eHealth intervention, should this be viewed as a consequence of the patient's ADHD or as a part of their personality? We therefore recommend that more studies apply the NASSS framework to evaluation of implementation and report on its applicability and suitability, which might lead to possible revisions or fine-tuning of the framework.

Finally, a strength of this study was the combination of different types of data to evaluate the implementation process. By using a mixed-methods approach, objective and more subjective data were combined, which proved to be valuable for gaining insight into a multilevel and elaborate implementation process. It was decided to not interview patients and management, since therapists were asked about the patient perspective and desk research provided insight into the strategies of management. Including these perspectives in interviews would probably not have produced much new information, but it is possible that factors were overlooked. Further research could focus on analyzing the patient perspective in implementation and investigate whether there are any discrepancies between therapist perspectives on patients and the patients' own perceptions. Finally, desk research was combined with interviews to provide a full picture of the implementation strategies, but not all information could be retrieved from the desk research (eg, number of participating therapists in the pilot). This shows the importance of carefully and fully documenting relevant information about implementation strategies from the start.

Conclusion

This study showed that the fit between the characteristics and needs of the therapists and patients, the organization, and the

technology was suboptimal, which has led to suboptimal implementation outcomes. An explanation for this could be the lack of a holistic approach in implementation: the implementation strategies mainly focused on training therapists' technical skills, while more attention should have been paid to necessary changes in the organization, an attitude change in therapists, and design of the technology. Here, adaptivity appears to be an important concept: a technology should be easily adaptable to an individual patient, therapists should be trained to be able to deal with an eMental health intervention in their treatment in a flexible way, and organizations must adapt their implementation strategies and structures to embed a new eHealth intervention. Consequently, in implementation, the holistic nature of eHealth and ensuring adaptivity on multiple levels appear to be pivotal.

Acknowledgments

Funding for this study was provided by Stichting Vrienden van Oldenkotte. We would like to thank Karen Rienks and Nathalie ten Cate, who conducted and transcribed the interviews, and Dirk Dijkslag for his help with the desk research and log data retrieval.

Conflicts of Interest

None declared.

References

1. Murray CJL, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2197-2223. [doi: 10.1016/S0140-6736(12)61689-4] [Medline: 23245608]
2. Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE, et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. Lancet 2013 Nov 9;382(9904):1575-1586. [doi: 10.1016/S0140-6736(13)61611-6] [Medline: 23993280]
3. Patel V, Chisholm D, Parikh R, Charlson FJ, Degenhardt L, Dua T, DCP MNS Author Group. Addressing the burden of mental, neurological, and substance use disorders: key messages from Disease Control Priorities, 3rd edition. Lancet 2016 Apr 16;387(10028):1672-1685. [doi: 10.1016/S0140-6736(15)00390-6] [Medline: 26454360]
4. Christensen H, Hickie IB. E-mental health: a new era in delivery of mental health services. Med J Aust 2010 Jun 07;192(11 Suppl):S2-S3. [Medline: 20528702]
5. Lindefors N, Andersson G. Guided Internet-Based Treatments in Psychiatry. Berlin: Springer; 2016.
6. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
7. Cuijpers P, Donker T, van Straten A, Li J, Andersson G. Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. Psychol Med 2010 Dec;40(12):1943-1957. [doi: 10.1017/S0033291710000772] [Medline: 20406528]
8. Schueller S. Mental health and eHealth technology. In: eHealth Research, Theory and Development. London: Routledge; 2018:111-130.
9. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019 Oct;18(3):325-336 [FREE Full text] [doi: 10.1002/wps.20673] [Medline: 31496095]
10. Valmaggia LR, Latif L, Kempton MJ, Rus-Calafell M. Virtual reality in the psychological treatment for mental health problems: an systematic review of recent evidence. Psychiatry Res 2016 Feb 28;236:189-195. [doi: 10.1016/j.psychres.2016.01.015] [Medline: 26795129]
11. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
12. Wentzel J, van der Vaart R, Bohlmeijer ET, van Gemert-Pijnen JEWC. Mixing online and face-to-face therapy: how to benefit from blended care in mental health care. JMIR Ment Health 2016;3(1):e9 [FREE Full text] [doi: 10.2196/mental.4534] [Medline: 26860537]
13. Lal S, Adair CE. E-mental health: a rapid review of the literature. Psychiatr Serv 2014 Jan 1;65(1):24-32. [doi: 10.1176/appi.ps.201300009] [Medline: 24081188]
14. van der Vaart R, Witting M, Riper H, Kooistra L, Bohlmeijer ET. Blending online therapy into regular face-to-face therapy for depression: content, ratio and preconditions according to patients and therapists using a Delphi study. BMC Psychiatry 2014;14(1):355 [FREE Full text] [doi: 10.1186/s12888-014-0355-z] [Medline: 25496393]
15. Gaggioli A. The bright future of technology in mental health. Cyberpsychol Behav Soc Networking 2018 Jun;21(6):399-400. [doi: 10.1089/cyber.2018.29115.csi]

16. Vis C, Mol M, Kleiboer A, Bührmann L, Finch T, Smit J, et al. Improving implementation of emental health for mood disorders in routine practice: systematic review of barriers and facilitating factors. JMIR Ment Health 2018 Mar 16;5(1):e20 [FREE Full text] [doi: 10.2196/mental.9769] [Medline: 29549072]
17. Powell BJ, McMillen JC, Proctor EK, Carpenter CR, Griffey RT, Bunger AC, et al. A compilation of strategies for implementing clinical innovations in health and mental health. Med Care Res Rev 2012 Apr;69(2):123-157 [FREE Full text] [doi: 10.1177/1077558711430690] [Medline: 22203646]
18. van Gemert-Pijnen L, Kelders M, Kip H, Sanderman R. eHealth Research, Theory and Development: A Multi-Disciplinary Approach. London: Routledge; 2018.
19. Waltz TJ, Powell BJ, Matthieu MM, Damschroder LJ, Chinman MJ, Smith JL, et al. Use of concept mapping to characterize relationships among implementation strategies and assess their feasibility and importance: results from the Expert Recommendations for Implementing Change (ERIC) study. Implement Sci 2015 Aug 07;10:109 [FREE Full text] [doi: 10.1186/s13012-015-0295-0] [Medline: 26249843]
20. Proctor EK, Landsverk J, Aarons G, Chambers D, Glisson C, Mittman B. Implementation research in mental health services: an emerging science with conceptual, methodological, and training challenges. Adm Policy Ment Health 2009 Jan;36(1):24-34 [FREE Full text] [doi: 10.1007/s10488-008-0197-4] [Medline: 19104929]
21. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
22. Emmelkamp PMG. Technological innovations in clinical assessment and psychotherapy. Psychother Psychosom 2005;74(6):336-343. [doi: 10.1159/000087780] [Medline: 16244509]
23. Varsi C, Solberg Nes L, Kristjansdottir OB, Kelders SM, Stenberg U, Zangi HA, et al. Implementation strategies to enhance the implementation of ehealth programs for patients with chronic illnesses: realist systematic review. J Med Internet Res 2019 Sep 27;21(9):e14255 [FREE Full text] [doi: 10.2196/14255] [Medline: 31573934]
24. Feijt MA, de Kort YA, Bongers IM, IJsselsteijn WA. Perceived drivers and barriers to the adoption of emental health by psychologists: the construction of the levels of adoption of emental health model. J Med Internet Res 2018 Dec 24;20(4):e153 [FREE Full text] [doi: 10.2196/jmir.9485] [Medline: 29691215]
25. 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. Lancet Psychiatry 2019 Mar;6(3):257-266. [doi: 10.1016/S2215-0366(18)30302-X] [Medline: 30522979]
26. Kivi M, Eriksson MC, Hange D, Petersson E, Björkelund C, Johansson B. Experiences and attitudes of primary care therapists in the implementation and use of internet-based treatment in Swedish primary care settings. Internet Interventions 2015 Sep;2(3):248-256. [doi: 10.1016/j.invent.2015.06.001]
27. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
28. van Gemert-Pijnen JEWC, Nijland N, Ossebaard HC, Kelders SM, Eysenbach G, Seydel ER. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
29. van Gemert-Pijnen L, Kip H, Kelders S, Sanderman R. Introducing eHealth. In: eHealth Research, Theory and Development. London: Routledge; 2018:23-46.
30. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
31. 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]
32. Pieterse M, Kip H, Martinez R. The complexity of eHealth implementation: a theoretical and practical perspective. In: eHealth Research, Theory and Development: A Multi-Disciplinary Approach. London: Routledge; 2018:247-270.
33. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q 2004;82(4):581-629 [FREE Full text] [doi: 10.1111/j.0887-378X.2004.00325.x] [Medline: 15595944]
34. Johnson RB, Onwuegbuzie AJ. Mixed methods research: a research paradigm whose time has come. Educ Res 2016 Jul;33(7):14-26. [doi: 10.3102/0013189X033007014]
35. Vis C, Kleiboer A, Prior R, Bønes E, Cavallo M, Clark SA, et al. Implementing and up-scaling evidence-based eMental health in Europe: the study protocol for the MasterMind project. Internet Interventions 2015 Nov;2(4):399-409. [doi: 10.1016/j.invent.2015.10.002]
36. Sieverink F, Kelders S, Poel M, van Gemert-Pijnen L. Opening the black box of electronic health: collecting, analyzing, and interpreting log data. JMIR Res Protoc 2017 Aug 07;6(8):e156 [FREE Full text] [doi: 10.2196/resprot.6452] [Medline: 28784592]

37. Venkatesh V, Brown S, Sullivan Y. Guidelines for conducting mixed-methods research: an extension and illustration. J Assoc Inf Syst 2016 Aug;17(7):435-494. [doi: 10.17705/1jais.00433]
38. Sieverink F, Kelders S, Braakman-Jansen A, van Gemert-Pijnen J. Evaluating the implementation of a personal health record for chronic primary and secondary care: a mixed methods approach. BMC Med Inform Decis Mak 2019 Nov 27;19(1):1. [doi: 10.1186/s12911-019-0969-7]
39. Mullen PE. Forensic mental health. Br J Psychiatry 2000 Apr;176:307-311. [doi: 10.1192/bjp.176.4.307] [Medline: 10827876]
40. Kip H, Bouman YHA, Kelders SM, van Gemert-Pijnen LJEWC. eHealth in treatment of offenders in forensic mental health: a review of the current state. Front Psychiatry 2018;9:42 [FREE Full text] [doi: 10.3389/fpsyt.2018.00042] [Medline: 29515468]
41. Creswell J, Clark V. Designing and Conducting Mixed Methods Research. Thousand Oaks: Sage Publications; 2017.
42. Greene J. Mixed Methods in Social Inquiry, Vol. 9. Hoboken: John Wiley & Sons; 2007.
43. Leech N, Onwuegbuzie A. Guidelines for conducting and reporting mixed research in the field of counseling and beyond. J Couns Devel 2010;88(1):61-69. [doi: 10.1002/j.1556-6678.2010.tb00151.x]
44. Nazi KM. The personal health record paradox: health care professionals' perspectives and the information ecology of personal health record systems in organizational and clinical settings. J Med Internet Res 2013;15(4):e70 [FREE Full text] [doi: 10.2196/jmir.2443] [Medline: 23557596]
45. Boeije H. A purposeful approach to the constant comparative method in the analysis of qualitative interviews. Qual Quant 2002;36(4):391-409. [doi: 10.4135/9781849209403.n920]
46. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Manag Sci 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
47. Venkatesh V, Morris M, Davis G, Davis F. User acceptance of information technology: toward a unified view. MIS Quarterly 2003;27(3):425. [doi: 10.2307/30036540]
48. van Velsen L, Evers M, Bara C, Op den Akker H, Boerema S, Hermens H. Understanding the acceptance of an ehealth technology in the early stages of development: an end-user walkthrough approach and two case studies. JMIR Form Res 2018 Jun 15;2(1):e10474 [FREE Full text] [doi: 10.2196/10474] [Medline: 30684434]
49. Legris P, Ingham J, Collerette P. Why do people use information technology? A critical review of the technology acceptance model. Inf Manag 2003 Jan;40(3):191-204. [doi: 10.1016/S0378-7206(01)00143-4]
50. Kip H. Technologie in de forensische zorg–Crossing borders. Kwaliteit Forensische Zorg 2019:1. [doi: 10.13140/RG.2.2.32303.69280]
51. Barakat A, Woolrych RD, Sixsmith A, Kearns WD, Kort HSM. eHealth technology competencies for health professionals working in home care to support older adults to age in place: outcomes of a two-day collaborative workshop. Med 2 0 2013;2(2):e10 [FREE Full text] [doi: 10.2196/med20.2711] [Medline: 25075233]
52. Bowen DE. The changing role of employees in service theory and practice: an interdisciplinary view. Hum Resource Manag Rev 2016 Mar;26(1):4-13. [doi: 10.1016/j.hrmr.2015.09.002]
53. Bierbooms J. Kwaliteit Forensische Zorg (KFZ). 2015. Do's en don'ts van e-health in de forensische ggz URL: https://tinyurl.com/y98w9xlh [accessed 2020-04-14]
54. Birgden A. Therapeutic jurisprudence and responsivity: finding the will and the way in offender rehabilitation. Psychol Crime Law 2004 Sep;10(3):283-295. [doi: 10.1080/10683160410001662771]
55. Levesque DA, Johnson JL, Welch CA, Prochaska JM, Fernandez AC. Computer-tailored intervention for juvenile offenders. J Soc Work Pract Addict 2012 Jan 1;12(4):391-411 [FREE Full text] [doi: 10.1080/1533256X.2012.728107] [Medline: 23264754]
56. Wild TSN, Fromberger P, Jordan K, Müller I, Müller JL. Web-based health services in forensic psychiatry: a review of the use of the internet in the treatment of child sexual abusers and child sexual exploitation material offenders. Front Psychiatry 2018;9:763 [FREE Full text] [doi: 10.3389/fpsyt.2018.00763] [Medline: 30778306]
57. Polaschek DL. Many sizes fit all: a preliminary framework for conceptualizing the development and provision of cognitive–behavioral rehabilitation programs for offenders. Aggr Violent Behav 2011 Jan;16(1):20-35. [doi: 10.1016/j.avb.2010.10.002]
58. Howner K, Andiné P, Bertilsson G, Hultcrantz M, Lindström E, Mowafi F, et al. Mapping systematic reviews on forensic psychiatric care: a systematic review identifying knowledge gaps. Front Psychiatry 2018;9:452 [FREE Full text] [doi: 10.3389/fpsyt.2018.00452] [Medline: 30319459]
59. Greenberg E, Dunleavy E, Kutner M. Literacy behind bars: results from the 2003 National Assessment of Adult Literacy Prison Survey. URL: https://nces.ed.gov/pubs2007/2007473.pdf [accessed 2020-04-14]
60. Dansky KH, Thompson D, Sanner T. A framework for evaluating eHealth research. Eval Program Plann 2006 Nov;29(4):397-404. [doi: 10.1016/j.evalprogplan.2006.08.009] [Medline: 17950868]
61. Abd-Alrazaq A, Bewick BM, Farragher T, Gardner P. Factors affecting patients' use of electronic personal health records in England: cross-sectional study. J Med Internet Res 2019 Jul 31;21(7):e12373 [FREE Full text] [doi: 10.2196/12373] [Medline: 31368442]
62. Smeijers D, Koole SL. Testing the effects of a virtual reality game for aggressive impulse management (VR-GAIME): study protocol. Front Psychiatry 2019;10:83 [FREE Full text] [doi: 10.3389/fpsyt.2019.00083] [Medline: 30863328]

63. Cornet L, Den Besten A, van Gelder J. Virtual reality en augmented reality in justitiële context. 2019. URL: https://ris.utwente.nl/ws/portalfiles/portal/107053177/Rapport_VRAR_in_justitie_context.pdf [accessed 2020-04-14]
64. Kip H, Kelders SM, Weerink K, Kuiper A, Brüninghoff I, Bouman YHA, et al. Identifying the added value of virtual reality for treatment in forensic mental health: a scenario-based, qualitative approach. Front Psychol 2019;10:406 [FREE Full text] [doi: 10.3389/fpsyg.2019.00406] [Medline: 30873093]
65. Klein Tuente S, Bogaerts S, van IJzendoorn S, Veling W. Effect of virtual reality aggression prevention training for forensic psychiatric patients (VRAPT): study protocol of a multi-center RCT. BMC Psychiatry 2018 Aug 06;18(1):251 [FREE Full text] [doi: 10.1186/s12888-018-1830-8] [Medline: 30081863]
66. Fromberger P, Jordan K, Müller JL. [Use of virtual reality in forensic psychiatry. A new paradigm?]. Nervenarzt 2014 Mar;85(3):298-303. [doi: 10.1007/s00115-013-3904-7] [Medline: 24549691]
67. de Looff P, Noordzij ML, Moerbeek M, Nijman H, Didden R, Embregts P. Changes in heart rate and skin conductance in the 30 min preceding aggressive behavior. Psychophysiology 2019 Oct;56(10):e13420. [doi: 10.1111/psyp.13420] [Medline: 31184379]
68. Cornet L, Mandersloot M, Pool R, de Kogel C. De 'zelfmetende'justitiabele. 2017. URL: https://www.wodc.nl/binaries/Cahier%202017-17_2716a_Volledige%20tekst_v1_tcm28-299893.pdf [accessed 2020-04-14]
69. Glasgow RE, Phillips SM, Sanchez MA. Implementation science approaches for integrating eHealth research into practice and policy. Int J Med Inform 2014 Jul;83(7):e1-e11. [doi: 10.1016/j.ijmedinf.2013.07.002] [Medline: 23910896]
70. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Comm Assoc Inf Syst 2009;24(1):28. [doi: 10.17705/1CAIS.02428]
71. Lehto T, Oinas-Kukkonen H, Drozd F. Factors affecting perceived persuasiveness of a behavior change support system. 2012 Presented at: Thirty-Third International Conference on Information Systems; 2012; Orlando.
72. Kelders SM, Kok RN, Ossebaard HC, van Gemert-Pijnen JEWC. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
73. Ludden GD, van Rompay TJ, Kelders SM, van Gemert-Pijnen JE. How to increase reach and adherence of web-based interventions: a design research viewpoint. J Med Internet Res 2015;17(7):e172 [FREE Full text] [doi: 10.2196/jmir.4201] [Medline: 26163456]
74. Asbjørnsen R, Smedsrød M, Solberg L, Wentzel J, Varsi C, Hjelmesæth J, et al. Persuasive system design principles and behavior change techniques to stimulate motivation and adherence in electronic health interventions to support weight loss maintenance: scoping review. J Med Internet Res 2019;21(6):e14265. [Medline: 31228174]
75. Beatty L, Binnion C. A systematic review of predictors of, and reasons for, adherence to online psychological interventions. Int J Behav Med 2016 Dec;23(6):776-794. [doi: 10.1007/s12529-016-9556-9] [Medline: 26957109]
76. van Limburg M, Wentzel J, Sanderman R, van Gemert-Pijnen L. Business modeling to implement an ehealth portal for infection control: a reflection on co-creation with stakeholders. JMIR Res Protoc 2015 Aug 13;4(3):e104 [FREE Full text] [doi: 10.2196/resprot.4519] [Medline: 26272510]
77. Potter J, Poulton R. Towards an evidence-informed plan of action for mental health and addiction in New Zealand. 2018. URL: https://tinyurl.com/y7mg7nj2 [accessed 2020-04-14]
78. Degl' Innocenti A, Hassing LB, Lindqvist A, Andersson H, Eriksson L, Hanson FH, et al. First report from the Swedish National Forensic Psychiatric Register (SNFPR). Int J Law Psychiatry 2014;37(3):231-237. [doi: 10.1016/j.ijlp.2013.11.013] [Medline: 24295538]
79. Goethals KR, Vorstenbosch EC, van Marle HJ. Diagnostic comorbidity in psychotic offenders and their criminal history: a review of the literature. Int J Forensic Mental Health 2008 Jan;7(2):147-156. [doi: 10.1080/14999013.2008.9914411]
80. Bloem O, Nijman H, Bulten B, Groen H, Drost M. Psychopathologie onder gedetineerden. In: Groen H, Drost M, Nijman H, editors. Handboek Forensische Geestelijke Gezondheidszorg. Utrecht: Uitgeverij De Tijdstroom; 2011:153-162.

Abbreviations

ADHD: attention deficit hyperactivity disorder
CFIR: consolidated framework for implementation research
eMental health: technology in mental health
NASSS: nonadoption, abandonment, scale-up, spread, and sustainability

Edited by G Eysenbach; submitted 04.11.19; peer-reviewed by C Varsi, H Ali; comments to author 12.12.19; revised version received 14.01.20; accepted 01.02.20; published 26.05.20.

Please cite as:
Kip H, Sieverink F, van Gemert-Pijnen LJEWC, Bouman YHA, Kelders SM
Integrating People, Context, and Technology in the Implementation of a Web-Based Intervention in Forensic Mental Health Care: Mixed-Methods Study
J Med Internet Res 2020;22(5):e16906
URL: http://www.jmir.org/2020/5/e16906/
doi:10.2196/16906
PMID:32348285

Original Paper

The Impact of a Tablet App on Adherence to American Heart Association Guidelines During Simulated Pediatric Cardiopulmonary Resuscitation: Randomized Controlled Trial

Johan N Siebert[1,2], MD; Laurence Lacroix[1], MD; Aymeric Cantais[3], MD; Sergio Manzano[1], MD; Frederic Ehrler[4], PhD

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

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

[3]Pediatric Emergency Department, University Hospital of Saint-Etienne, Saint-Etienne, France

[4]Diagnostic Department, Geneva University Hospitals, Geneva, Switzerland

Corresponding Author:
Johan N Siebert, MD
Department of Pediatric Emergency Medicine
Geneva Children's Hospital
Geneva University Hospitals
Avenue de la Roseraie, 47
1211 Geneva 14
Switzerland
Phone: 41 795534072
Fax: 41 22 372 54 23
Email: Johan.Siebert@hcuge.ch

Abstract

Background: Evidence-based best practices are the cornerstone to guide optimal cardiopulmonary arrest resuscitation care. Adherence to the American Heart Association (AHA) guidelines for cardiopulmonary resuscitation (CPR) optimizes the management of critically ill patients and increases their chances of survival after cardiac arrest. Despite advances in resuscitation science and survival improvement over the last decades, only approximately 38% of children survive to hospital discharge after in-hospital cardiac arrest and only 6%-20% after out-of-hospital cardiac arrest.

Objective: We investigated whether a mobile app developed as a guide to support and drive CPR providers in real time through interactive pediatric advanced life support (PALS) algorithms would increase adherence to AHA guidelines and reduce the time to initiation of critical life-saving maneuvers compared to the use of PALS pocket reference cards.

Methods: This study was a randomized controlled trial conducted during a simulation-based pediatric cardiac arrest scenario caused by pulseless ventricular tachycardia (pVT). A total of 26 pediatric residents were randomized into two groups. The primary outcome was the elapsed time in seconds in each allocation group from the onset of pVT to the first defibrillation attempt. Secondary outcomes were time elapsed to (1) initiation of chest compression, (2) subsequent defibrillation attempts, and (3) administration of drugs, including the time intervals between defibrillation attempts and drug doses, shock doses, and the number of shocks. All outcomes were assessed for deviation from AHA guidelines.

Results: Mean time to the first defibrillation attempt (121.4 sec, 95% CI 105.3-137.5) was significantly reduced among residents using the app compared to those using PALS pocket cards (211.5 sec, 95% CI 162.5-260.6, *P*<.001). With the app, 11 out of 13 (85%) residents initiated chest compressions within 60 seconds from the onset of pVT and 12 out of 13 (92%) successfully defibrillated within 180 seconds. Time to all other defibrillation attempts was reduced with the app. Adherence to the 2018 AHA pVT algorithm improved by approximately 70% (*P*=.001) when using the app following all CPR sequences of action in a stepwise fashion until return of spontaneous circulation. The pVT rhythm was recognized correctly in 51 out of 52 (98%) opportunities among residents using the app compared to only 19 out of 52 (37%) among those using PALS cards (*P*<.001). Time to epinephrine injection was similar. Among a total of 78 opportunities, incorrect shock or drug doses occurred in 14% (11/78) of cases among those using the cards. These errors were reduced to 1% (1/78, *P*=.005) when using the app.

Conclusions: Use of the mobile app was associated with a shorter time to first and subsequent defibrillation attempts, fewer medication and defibrillation dose errors, and improved adherence to AHA recommendations compared with the use of PALS pocket cards.

(J Med Internet Res 2020;22(5):e17792) doi:10.2196/17792

KEYWORDS

biomedical technologies; mobile apps; emergency medicine; pediatrics; resuscitation; guideline adherence

Introduction

Pediatric cardiac arrest is a high-risk, low-frequency event associated with death or severe neurological sequelae in survivors. It requires immediate recognition and care by skilled health providers. Recent studies show that pediatric in-hospital cardiac arrest (p-IHCA) affects 7100-8300 children per year in the United States [1], of which 14% occur in pediatric emergency departments (PEDs) [2]. Pediatric out-of-hospital cardiac arrest (p-OHCA) accounts for a further 7037 children brought to US PEDs by emergency medical services each year [3]. Despite advances in resuscitation science and survival improvement over the last decades, survival remains low, with only approximately 38% of children surviving to hospital discharge after p-IHCA, and 6%-20% after p-OHCA [3,4]. Evidence-based best practices are the cornerstone for the guidance of optimal cardiopulmonary arrest resuscitation care. High-quality cardiopulmonary resuscitation (CPR), according to the American Heart Association (AHA) life-support guidelines, is associated with a successful return of spontaneous circulation (ROSC), improved survival after hospital discharge, and good neurological outcomes [5]. Deviation from recommended procedures is associated with a reduced likelihood of survival from cardiac arrest [6].

While adherence to AHA guidelines in emergency departments has been described for adults, there are limited data for PEDs [7]. Reference tools for pediatric emergency physicians to handle pediatric CPR according to AHA guidelines are available on reference pocket cards. Unfortunately, health care providers frequently do not perform resuscitation according to guidelines, despite cognitive aids [8] and AHA life-support training courses, such as basic life support (BLS) and pediatric advanced life support (PALS). Suboptimal quality of CPR is still commonly encountered for both adult and pediatric patients [9].

New resuscitation strategies using information technologies and devices aiming to improve both in- and out-of-hospital CPR have been assessed to ensure adherence to AHA guidelines [10-16]. Nevertheless, research in this area remains scarce, especially in pediatrics, and studies assessing the impact of information technology on p-IHCA management and improved pediatric CPR outcomes are necessary. In a previous randomized trial, we found that adherence to PALS algorithms when adapted on Google Glass was improved with a significant reduction of errors and deviations in defibrillation doses by 53% when compared to the use of pocket reference cards [17]. However, time to the first defibrillation attempt and adherence to AHA guidelines to other critical resuscitation endpoints in terms of time and drug-dose delivery were not improved using the glasses. The complexity of interacting while wearing glasses, as well as the limits of the system to situate the current action in the whole resuscitation process and their small size, were major limitations to their potential use in p-IHCA. Thus, we have developed a new mobile app—the Guiding Pad app—from the ground up and dedicated it to tablets. It is intended as a guide to support and drive CPR providers in real-time conditions through interactive PALS algorithms enhanced with patient-centered cognitive aids.

Our objective was to investigate, in a simulated model, whether this app would increase adherence to AHA guidelines by reducing deviation and time to initiation of critical life-saving maneuvers during pediatric CPR compared to the use of PALS pocket reference cards.

Methods

Study Design

We conducted a prospective, randomized controlled trial in a tertiary PED (>33,000 consultations/year) with two parallel groups of voluntary pediatric residents. We compared time to the first defibrillation attempt and other critical resuscitation endpoints using a tablet app (Guiding Pad, group A) or AHA PALS conventional pocket reference cards (group B) during a standardized simulation-based pediatric cardiac arrest scenario. No changes were made to the app during the study.

The trial received a declaration of *no objection* by Swissethics and the Geneva Cantonal Ethics Committee, as the purpose of the study was to examine the effect of the intervention on health care providers. For the same reason, and according to the International Committee of Medical Journal Editors, a trial registration number was not required. The trial was conducted according to the principles of the Declaration of Helsinki and Good Clinical Practice guidelines, and in accordance with the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online TeleHealth) guidelines (see Multimedia Appendix 1) [18] and the Reporting Guidelines for Health Care Simulation Research [19].

Participants

Any physician performing a residency in pediatrics was eligible. Shift-working residents were randomly recruited on the day of the study using an alphabetical list to avoid preparation bias. Included participants benefited from a standardized 5-minute introduction course on the use of the tablet app. As BLS training is a requirement for residents at our institution, all participants had previously completed this course prior to study entry. Participation in a simulation in the past month was an exclusion criterion to avoid a recent training effect. Study participants

were not involved in the study design, choice of outcome measures, or the execution of the study. No participant was asked for advice on the interpretation or writing of the study results. Participants were informed of the results after completion of the study.

Randomization and Blinding

Residents were randomized using a single constant 1:1 allocation ratio determined with the web-based randomization software Sealed Envelope [20]. Written informed consent was obtained from each participant after full information disclosure prior to study participation. Blinding to the purpose of the study was maintained during recruitment to minimize preparation bias. Allocation concealment was managed with the software and was not released until the participant started the scenario.

The Guiding Pad App

Unlike adults, cardiac arrest in children without prior cardiac disease is mainly due to asystole (40%) and pulseless electrical activity (24%) [2]. As ventricular fibrillation and pulseless ventricular tachycardia (pVT), namely shockable rhythms, have been identified in 27% of p-IHCA cases [21], we decided to use the pVT algorithm, as we considered that it would offer a greater opportunity to assess the multiple-step resuscitative skills set out in the AHA guidelines. The app was developed at Geneva University Hospitals using Angular, version 8, a development framework created by Google to build mobile and web apps. The AHA PALS algorithms were adapted for tablets following a user-centered and ergonomic-driven approach by computer scientists, senior pediatric emergency physicians, and ergonomists. The numerous steps of AHA PALS algorithms were split into stages. Each stage transposed to the tablet paralleled the informational content of a resuscitation step from the original algorithm. The algorithms thus obtained were set up in a manner similar to the PALS pocket references regarding the progression and sequence of actions along the original algorithms' sequences. For instance, the complete pVT algorithm was designed to be as concise as possible without hindering proper progression along the algorithm. After completing a quick pediatric assessment triangle as the first step to recognize cardiac arrest and initiate CPR, including information on the weight or age of the patient, the app displays an initial statement about whether the pulse is present or not and the subsequent cardiac arrhythmias. Once selected, the screen then splits into two main sections.

The first section on the left-hand side of the screen displays an entire algorithm overview, thus allowing users to situate the stepwise resuscitation progress in real time along that algorithm. The current step of the resuscitation process is surrounded by a blinking red line that allows an immediate understanding of the current position within the algorithm. Each action already performed turns grey. However, the app provides the possibility for users to navigate back and forth through the algorithm at any time in order to select one of the resuscitation steps if needed.

The second section on the right-hand side of the screen displays the following elements:

1. A color-coded title allowing direct identification of each step in progress.
2. Cognitive aids helping with decision making, such as a distinctive illustration of cardiac rhythms (see Figure 1) or the shock dose to deliver with a picture of a manual defibrillator (Philips HeartStart MRx Biphasic Defibrillator, Philips Medical Systems) (see Figure 2).
3. A detailed and clickable list of actions to perform in a stepwise manner.
4. A footer to preview the next step.

Each action prompts the provider either to perform a choice (ie, choose the correct arrhythmia among several propositions) or to validate an action to be executed (eg, a drug-dose administration), which is brought to the attention of the provider by a red-box warning (see Figure 3). Weight-based drug doses are automatically calculated by an in-built engine already used in another evidence-based app that was assessed in a multicenter randomized controlled trial for in-hospital emergency drug delivery [22]. Shock doses using Philips HeartStart MRx Biphasic Defibrillator are automatically calculated based on patients' weights or ages. Each cycle of chest compression-ventilation is timed by a countdown clock displayed on the screen. In the case of rhythm change, the user can easily navigate across the multiple PALS algorithms (bradycardia, supraventricular tachycardia with poor perfusion, etc) at any time. All actions performed by the provider are automatically saved in log files to preserve information that can be retrieved at any time for debriefing or medicolegal purposes.

Figure 1. A screenshot of the Guiding Pad app. The left-hand side of the screen displays the American Heart Association (AHA) pediatric advanced life support (PALS) ventricular fibrillation (VF) and pulseless ventricular tachycardia (pVT) cardiac arrest algorithm. The current step (eg, determining the shockable status of the arrhythmia) of the resuscitation process is surrounded by a blinking red line. Past actions already accomplished are shown as shaded. At the top right-hand side of the screen, a color-coded title depicts the current step in progress. Below, four pulseless dysrhythmias are displayed; the provider selects the right one under consideration. At the bottom right-hand side, a footer helps to anticipate the next cardiopulmonary resuscitation (CPR) step. IO: intraosseous; IV: intravenous.

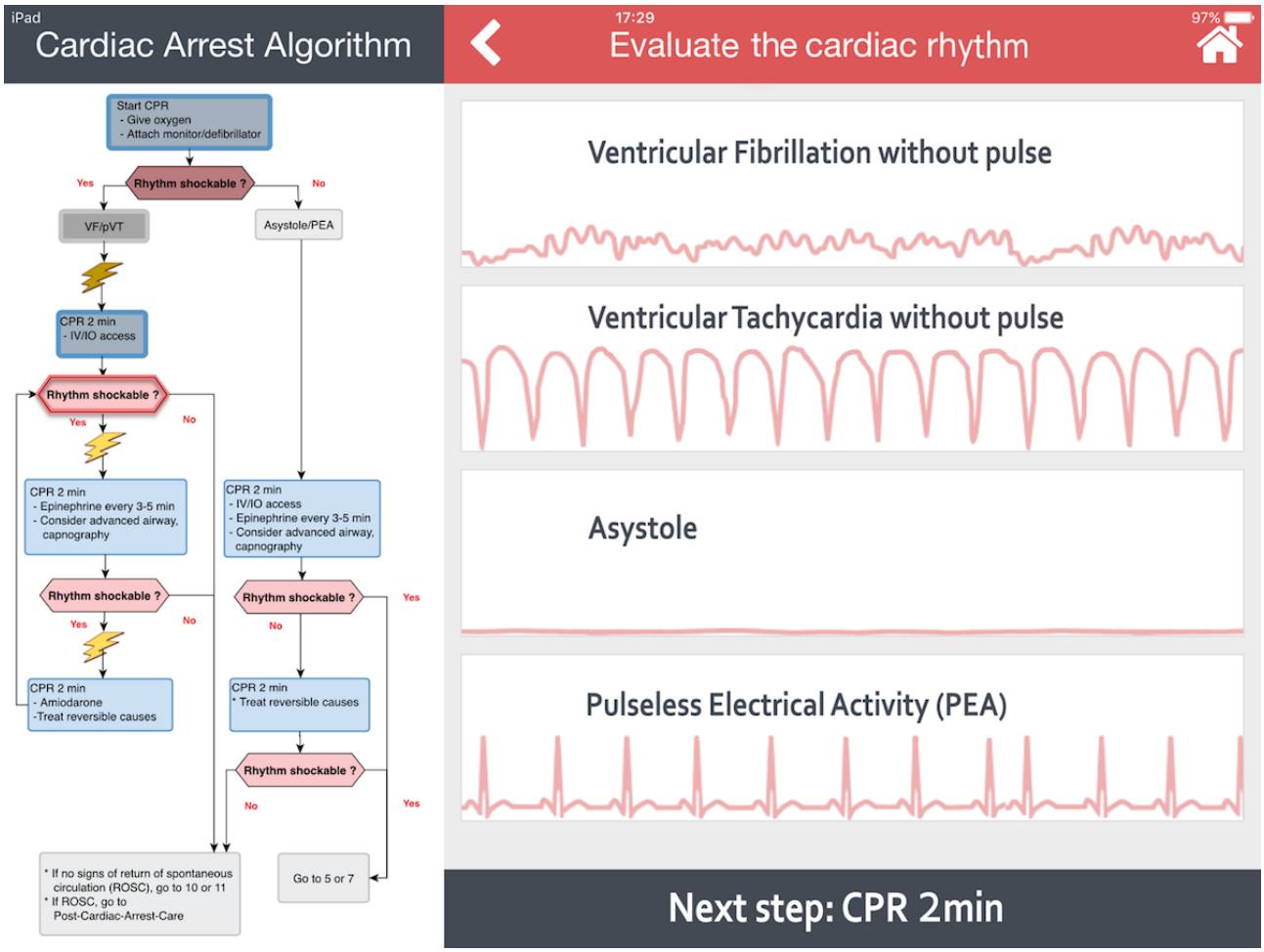


Figure 2. Screenshot of the Guiding Pad app. The left-hand side of the screen displays the American Heart Association (AHA) pediatric advanced life support (PALS) ventricular fibrillation (VF) and pulseless ventricular tachycardia (pVT) cardiac arrest algorithm. The current step (eg, defibrillation) of the resuscitation process is displayed with a red lightning bolt. Past actions already accomplished are shown as shaded. On the right-hand side of the screen, the weight-based shock dose to deliver is displayed with a picture of a manual defibrillator (Philips HeartStart MRx Biphasic Defibrillator). Once delivered, clicking the "Shock delivered" button validates the action and allows the user to proceed to the next step. At the bottom right-hand side, a footer helps to anticipate the next compression:ventilation step. CPR: cardiopulmonary resuscitation; IO: intraosseous; IV: intravenous; PEA: pulseless electrical activity.

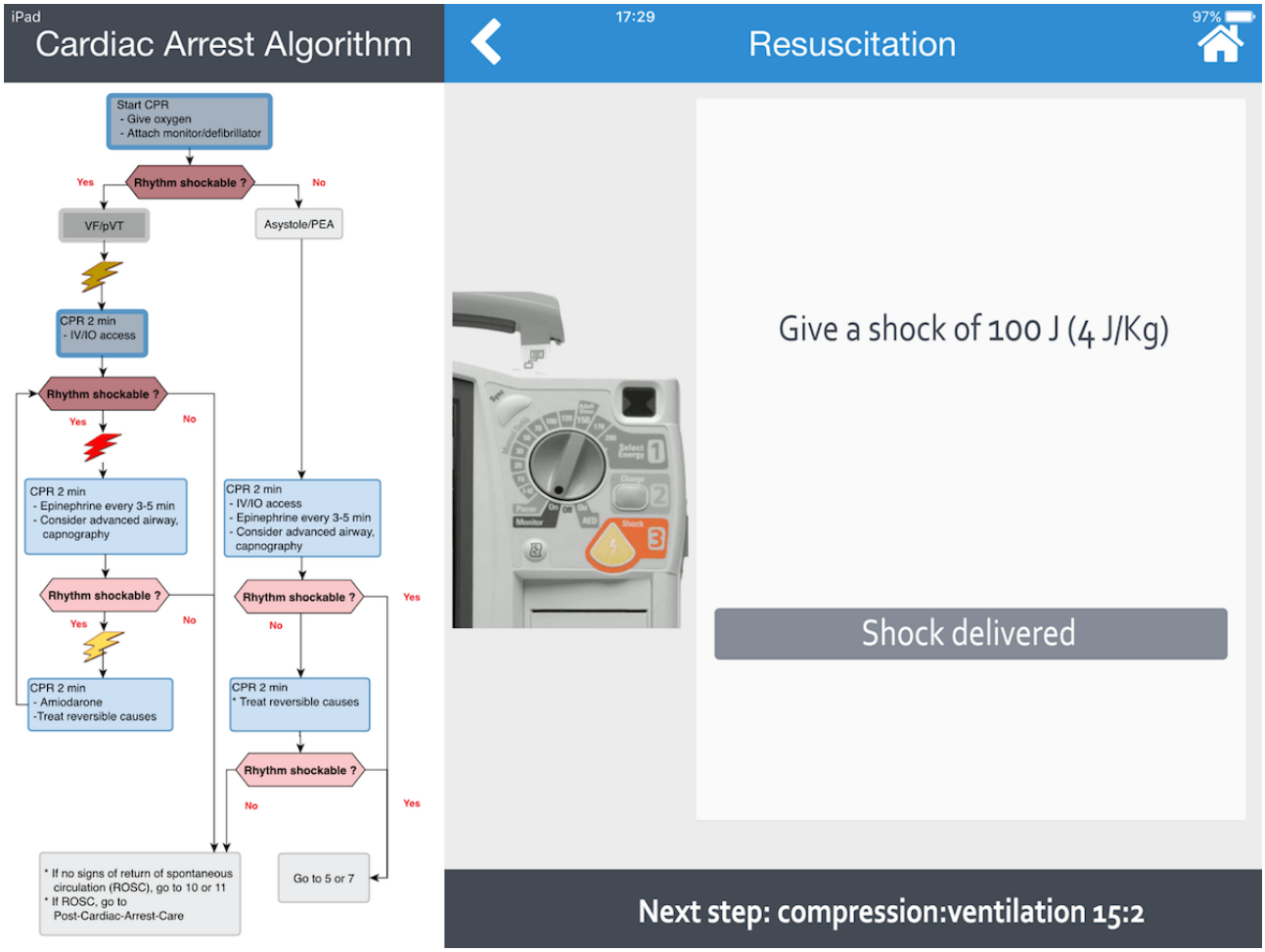
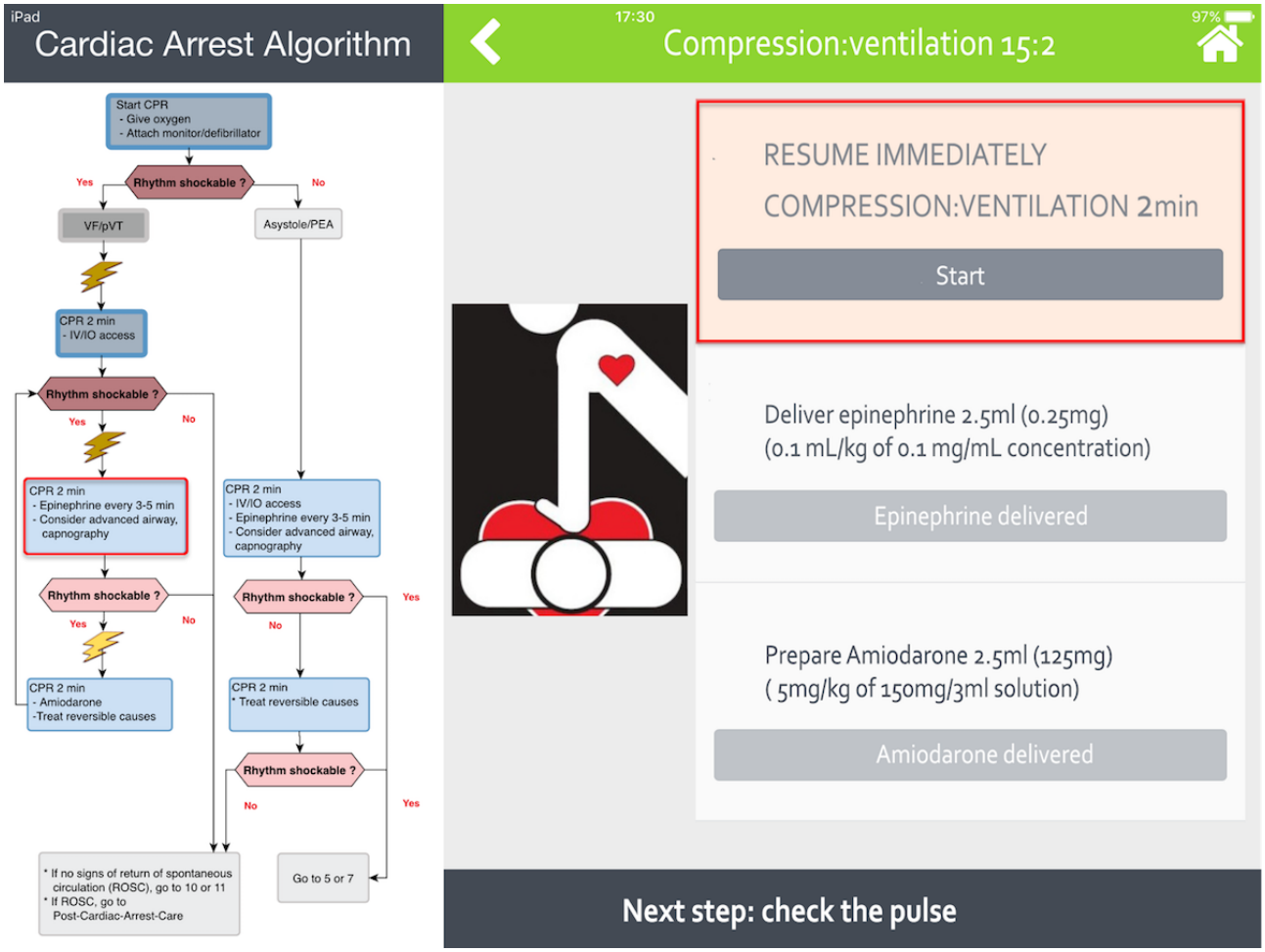


Figure 3. Screenshot of the Guiding Pad app. The left-hand side of the screen displays the American Heart Association (AHA) pediatric advanced life support (PALS) ventricular fibrillation (VF) and pulseless ventricular tachycardia (pVT) cardiac arrest algorithm. On the right-hand side of the screen, the sequence of actions to be taken are displayed in a stepwise manner to facilitate accurate progression along the algorithm. The current action (eg, to resume compression and ventilation) is brought to the attention of the provider by a red-box warning and requires validation by a simple click. Once completed, the next action will be to deliver the weight-based epinephrine dose automatically calculated by the app and then to prepare amiodarone. The next step shown at the bottom right-hand side will be to check the pulse. CPR: cardiopulmonary resuscitation; IO: intraosseous; IV: intravenous; PEA: pulseless electrical activity.



Procedures

On the day of participation, each resident completed an anonymous survey on basic demographic information, professional length of clinical experience, and PALS training. After random allocation, each participant received a standardized 5-minute training session on how to use the app. Participants were then asked to perform a 15-minute, highly realistic, scripted CPR scenario on a high-fidelity manikin (SimJunior; Laerdal Medical). The scenario was standardized to strictly follow the 2018 AHA pediatric pVT algorithm (see Figure 4) and was performed on the same high-fidelity manikin already primed with vital signs appropriate for the scenario (see Multimedia Appendix 2). It was conducted in situ in the PED shock room to increase realism, thus allowing participants to make use of real resources in the actual environment where they were expected to handle cardiac arrest. All participants in group B were offered the possibility to hold PALS pocket reference cards in their hands throughout the entire scenario. Whether they referred to them or not was left to their discretion, similar to real-life settings. No interactions occurred between participants and investigators. The simulation involved the participating resident and a resuscitation team comprising three study team members (ie, a PED registered nurse and two medical students) to assist with resuscitation through drug preparation, chest compression, and bag-valve-mask ventilation. Study team members had no role in decision making to achieve ROSC. A PALS instructor (ie, a pediatric emergency physician) who was not a member of the resuscitation team operated the simulator. To be consistent with the 2018 AHA pediatric cardiac arrest algorithm [23] and to standardize the scenario, defibrillation doses of 2 Joules per kg for the first attempt, and 4 Joules per kg for the subsequent second, third, and fourth attempts, were expected (see Figure 4).

Epinephrine and amiodarone drug doses had to be given just before or just after the second or third shock attempts, respectively (see Figure 5). The ROSC as demonstrated by a palpable pulse and signs of regaining consciousness corresponded to the end of the scenario.

Figure 4. American Heart Association (AHA) pediatric cardiac arrest algorithm: 2018 update (Duff et al, 2018). CPR: cardiopulmonary resuscitation; ET: endotracheal tube; IO: intraosseous; IV: intravenous; PEA: pulseless electrical activity; pVT: pulseless ventricular tachycardia; VF: ventricular fibrillation.

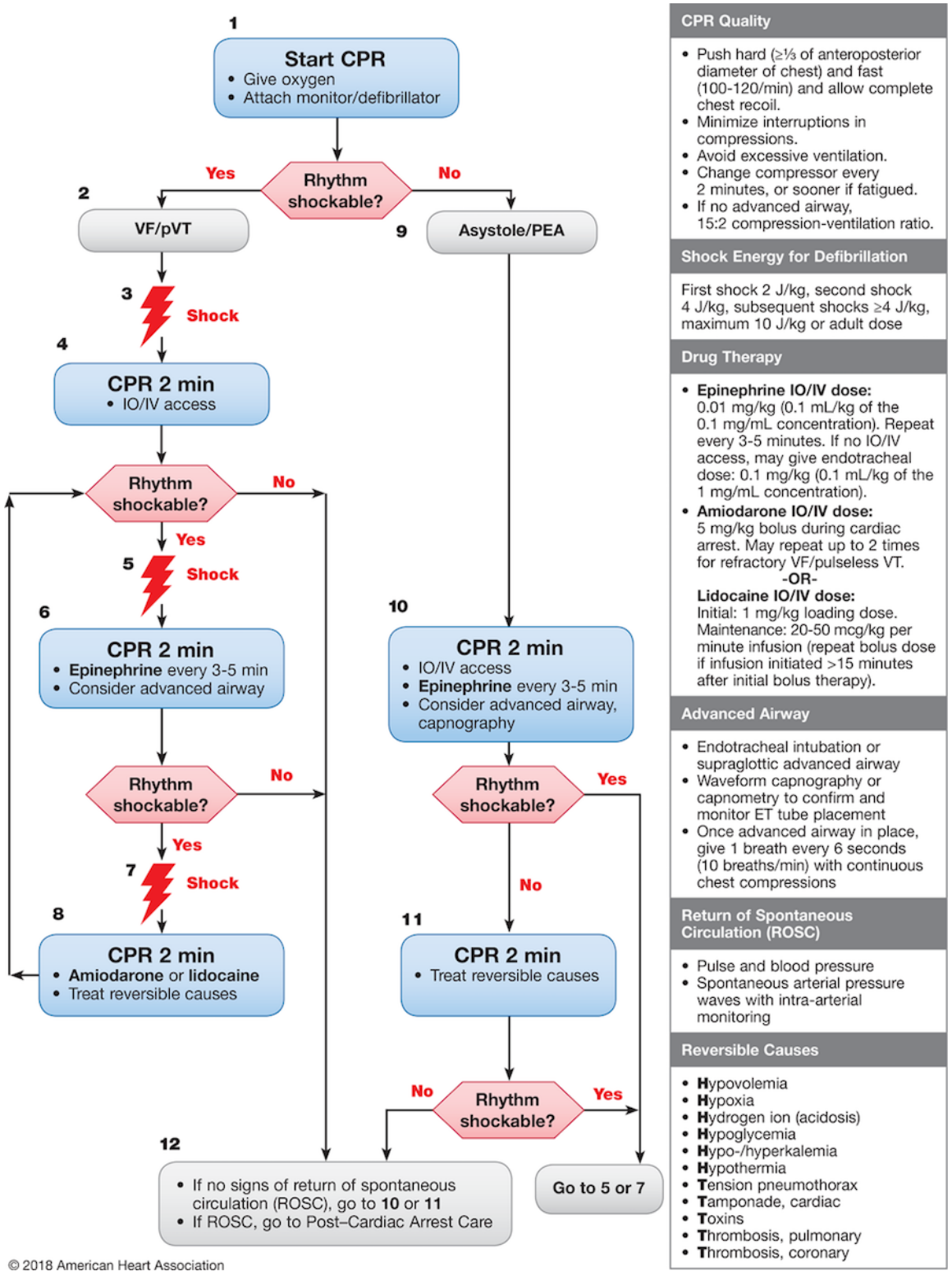
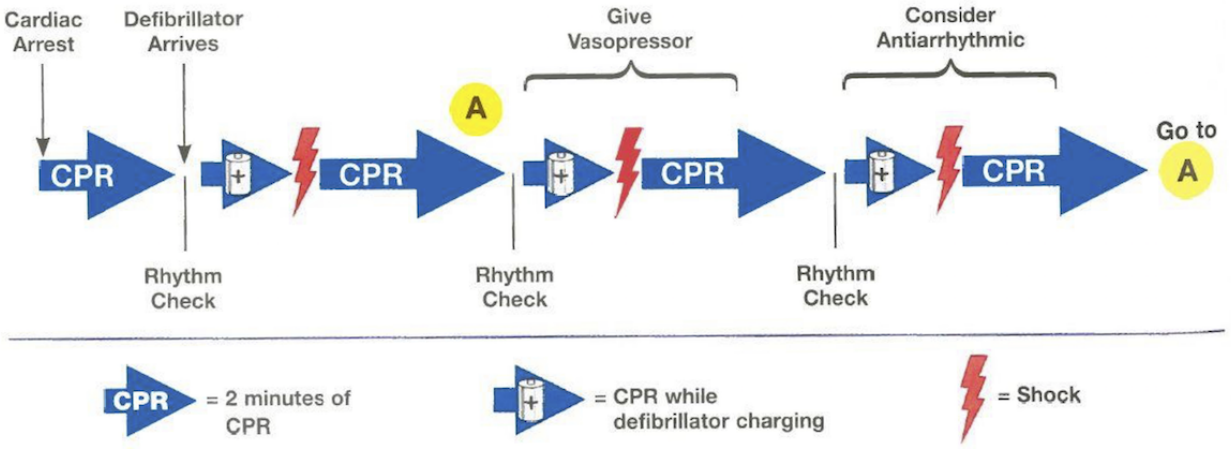


Figure 5. Summary of the ventricular fibrillation and pulseless ventricular tachycardia (pVT) cardiac arrest sequence. This original illustration is from the eBook edition of the Pediatric Advanced Life Support (PALS) Instructor Manual, published by the American Heart Association (AHA), 2015. CPR: cardiopulmonary resuscitation.



Outcomes

The primary outcome was the delay (in seconds) in each allocation group from the end of the clinical statement given by the study investigator to the first defibrillation attempt, as the expected survival advantage from early CPR can be significantly affected by a subsequent delay in defibrillation [24,25]. Secondary outcomes were the delay (in seconds) to initiation of chest compression; time to subsequent defibrillation attempts; time to administration of epinephrine and amiodarone; time interval (in seconds) between defibrillation attempts, drug doses, shock doses, and number of shocks; and perceived stress and satisfaction scores after completion of the scenario, as measured by a questionnaire using 10-point Likert scales (see Multimedia Appendix 3). The AHA recommends five cycles of chest compression (approximately 2 minutes) between each defibrillation attempt. The time spent by participants to perform chest compressions by compression cycles was defined as the hands-on time and was measured in seconds with a chronometer. All these outcomes were assessed for deviation from AHA guidelines.

Methods of Measurement and Data Collection

All actions (ie, primary and secondary outcomes) performed by the resident during the scenario were independently recorded by two trained investigators blinded to each other's records during the simulation, thus allowing an accurate assessment of timing and sequencing of actions and avoiding assessment bias. In the case of disagreement, a third independent evaluator helped reach a consensus. Data were manually retrieved and entered into a Microsoft Excel spreadsheet, version 16 (Microsoft Corporation). Unaccomplished actions were left blank and time was not assigned. Residents' privacy was preserved. Only the study investigators had access to the data.

Statistical Analysis

Power calculations were based on the detection of a 30-second decrease in time to the first defibrillation attempt between the two independent groups. A previous study has shown a mean time to first defibrillation of 92 seconds with an SD of 23 seconds [26]. Assuming a similar SD in each group of our study, 10 participants per group had to be recruited to provide the trial with 80% power at a two-sided alpha level of .05. To prevent a potential loss of power due to misspecification of assumptions, 13 participants were recruited per group, giving a total sample size of 26 participants.

For the primary analysis, we first evaluated the time elapsed between the onset of pVT and first defibrillation attempt. The Shapiro-Wilks test was used for normality analysis of the parameters. As most of the continuous variables were normally distributed, means and SDs with their 95% CIs were reported. Nonnormally distributed variables were analyzed using a Mann-Whitney test. Frequencies were reported as percentages. We used *t* tests to compare independent groups. No paired data were compared. Kaplan-Meier curves for time elapsed between the onset of pVT and first defibrillation attempt were estimated and compared using the log-rank (Mantel-Cox) test for bivariate survival analysis.

For the secondary analysis, we evaluated the time elapsed between the onset of pVT to subsequent defibrillation attempts and the delivery of both drugs. For normally distributed variables, means and SDs with 95% CI were reported. Nonnormally distributed variables were analyzed using a Mann-Whitney test. Frequencies were reported as percentages. We used *t* tests to compare independent groups. No paired data were compared. Kaplan-Meier curves for time elapsed between the onset of pVT and subsequent defibrillation attempts and delivery of both drugs were also estimated and compared using the log-rank (Mantel-Cox) test for bivariate survival analysis. Errors in cycles of chest compression-ventilation were measured as the deviation in percent from the experimental time spent in seconds compared to the 2-minute duration recommended by the AHA. Incorrect defibrillation or drug doses were measured as a deviation from the amount of energy delivered in Joules or drug doses in milliliters compared to AHA recommendations. A chi-square test was used to assess the relationship between absolute errors in defibrillation and drug doses expressed as categorical variables. Incorrect defibrillation mode was also measured. Absolute deviations were also analyzed. The mean

(SD) difference in deviation obtained with each method was reported with a 95% CI. A *t* test for unpaired data was used to compare interventions. Mean differences were reported by randomized group. Univariate linear regression analyses with 95% CI were performed to assess whether time to initiation of chest compression, defibrillation attempts, and drug delivery were associated with the number of postgraduate years or prior resuscitation experience as a provider in real-life and simulated environments. Means and SDs were determined for the perceived stress and satisfaction scores of individuals derived from the Likert-scale questionnaire and reported with descriptive statistics. A *P* value less than .05 was considered significant.

Interrater reliability was assessed by two observers who independently evaluated each resident's performance. Interrater reliability scores were calculated using the Cohen kappa coefficient for the shock and drug-dose errors. As the remaining outcomes were continuous variables, the Bland-Altman method was used to plot the difference of values reported by both observers against the mean value for each outcome. The limits of agreement were assessed by the interval of SD 1.96 of the measurement differences on either side of the mean difference. The null hypothesis that there was no difference, on average, between both reviewers was tested using a *t* test. The mean difference was reported with its 95% CI. Additionally, the intraclass correlation coefficients for times to each critical endpoint were assessed, assuming that raters were comprised of a sample from a larger population of possible raters. GraphPad Prism, version 8 (GraphPad Software), and SPSS, version 15.0 (SPSS Inc), were used for graph figures and to perform descriptive and statistical analyses.

Results

Study Participants

From August 30 to October 17, 2019, 26 pediatric residents were assessed for eligibility and randomly assigned to either the Guiding Pad app group (group A) (n=13) or the PALS pocket card group (group B) (n=13), without any dropouts or missing data (see Figure 6). Baseline characteristics were balanced in the two groups (see Table 1). In particular, we observed no statistically significant difference between the ages of both randomization arms (ie, no bias in randomization). We observed perfect interrater agreement for the scoring of the pVT scenario (see Table S1 in Multimedia Appendix 4, Table S2 in Multimedia Appendix 5, and Figure S1 in Multimedia Appendix 6).

Figure 6. Trial flowchart. PALS: pediatric advanced life support.

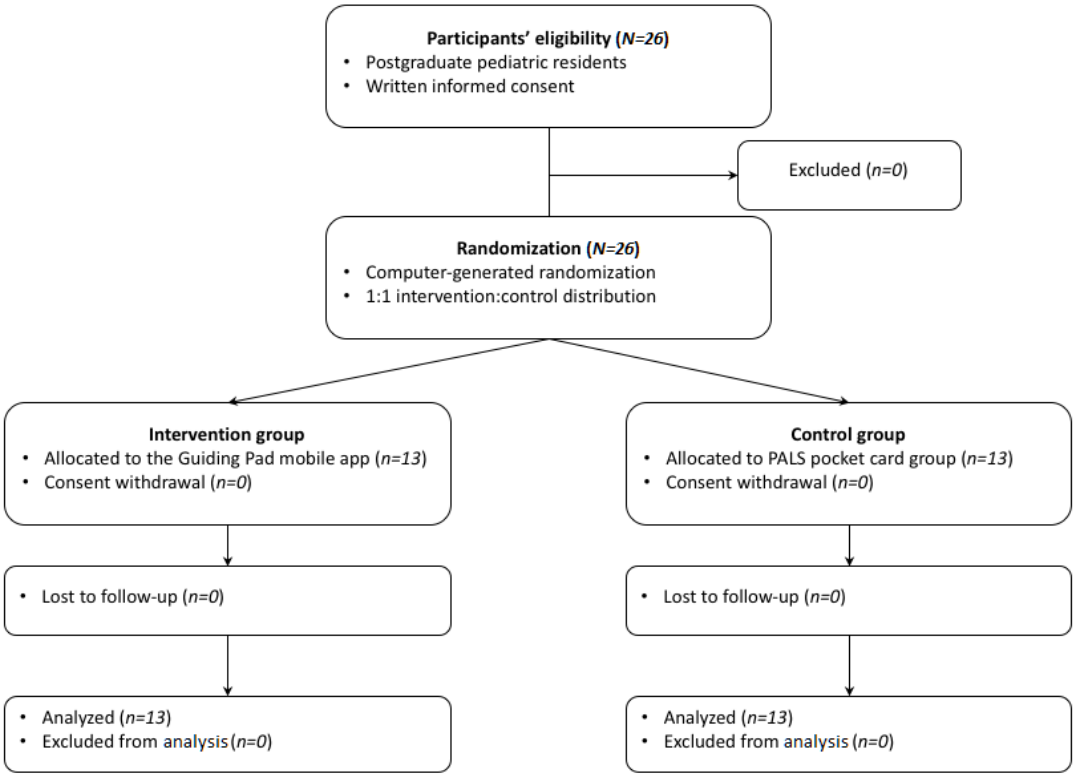


Table 1. Participants' demographics and clinical characteristics.

Demographics and clinical characteristics	Randomization arm	
	Guiding Pad (n=13)	PALS[a] pocket cards (n=13)
Age in years[b], mean (SD)	31.2 (5.5)	29.0 (2.2)
Sex (female), n (%)	9 (69)	10 (77)
Years of residency, mean (SD)	3.7 (2.3)	3.5 (0.9)
Number of basic life support providers among residents, n (%)	13 (100)	13 (100)
Number of PALS providers among residents, n (%)	9 (69)	11 (85)
Level of self-confidence[c] in following American Heart Association guidelines, mean (SD)	2.9 (1.1)	3.5 (0.7)
Number of residents having been enrolled in more than five resuscitations in the past, n (%)	4 (31)	7 (54)
Prior simulation-based resuscitations, mean (SD); total	6.5 (6.2); 84	4.8 (3.5); 62
Prior real-world cardiopulmonary resuscitations, mean (SD); total	9.2 (21.4); 120	7.8 (7.3); 102
Prior use of a manual defibrillator in either real-world or simulated environments, n (%)	10 (77)	9 (69)
Months since last manual-mode defibrillator use in either real-world or simulated environments, mean (SD)	9.5 (10.2)	13.9 (17.8)

[a]PALS: pediatric advanced life support.

[b]The age difference between the randomization arms was not statistically significant.

[c]Self-confidence was measured on a scale of 1 (not confident) to 5 (very confident).

Time to Critical Resuscitation Endpoint

Using the Guiding Pad app, 11 residents out of 13 (85%) initiated chest compressions within 60 seconds of the onset of pVT (9/13, 69%, within 30 sec), and 12 out of 13 (92%) successfully defibrillated within 180 seconds (see Figure 7). Mean time elapsed between the onset of pVT and first defibrillation attempt was 121.4 seconds (SD 26.7).

With PALS pocket cards, out of 13 residents, 10 (77%) started compressions within 60 seconds, 1 (8%) started compressions 277 seconds after onset of pVT, and 6 (46%) failed to discharge the defibrillator within 180 seconds. Mean time from initiation of chest compression to the first shock was significantly reduced for residents using the app (89.3 sec) than for those using the PALS pocket cards (163 sec; *P*=.002). Mean times to other critical resuscitation endpoints are summarized in Table 2. All defibrillation attempts, as well as amiodarone administration, were delivered significantly earlier in group A than in group B. However, the app was unable to speed up the delay before intraosseous access and epinephrine delivery (see Table 2 and Figure 8). We sought to analyze, in both groups, the difference in mean time to first defibrillation attempts between residents with or without previous defibrillation experience in either real-world or simulated environments, but with our small sample size we did not find any difference (with the app: 124.1 vs 112.3 sec, *P*=.53; without the app: 268.8 vs 186.1 sec, *P*=.09).

At the time of the study, 24 participants out of 26 (92%) were residents with more than one year of pediatric training (ie, postgraduate years). In a simple linear regression model, using the app was associated with a significant or borderline significant reduction in time to defibrillation attempts, regardless of the postgraduate years, and less scattered delays around the mean defibrillation time than when using the pocket cards (see Figure 9).

In group B, time to defibrillation attempts was inversely associated with the number of postgraduate years. In both groups, we observed no correlation between the time to initiation of chest compression or time to drug delivery and postgraduate years (see Figure 10). Moreover, we observed no relationship between previous CPR experience expressed as the number of prior CPR attempts on either a patient or a manikin and times to initiation of CPR, defibrillation attempts, or drug delivery (see Figure S2 in Multimedia Appendix 7 and Figure S3 in Multimedia Appendix 8).

Figure 7. Time to cardiopulmonary resuscitation (CPR) and defibrillation attempts. Kaplan-Meier curves of time elapsed between the onset of simulated pulseless ventricular tachycardia (pVT) and initiation of chest compression (ie, CPR) for the first, second, third, and fourth defibrillation attempts for residents using the Guiding Pad (GP) app vs conventional pediatric advanced life support (PALS) pocket cards (PCs). Log-rank (Mantel-Cox) test comparing curves: χ^2=0.0 and *P*=.97 for CPR; χ^2=13.9 and *P*<.001 for the first defibrillation attempt; χ^2=8.9 and *P*=.003 for the second defibrillation attempt; χ^2=13.3 and *P*<.001 for the third defibrillation attempt; and χ^2=9.5 and *P*=.002 for the fourth defibrillation attempt. ns: not significant.

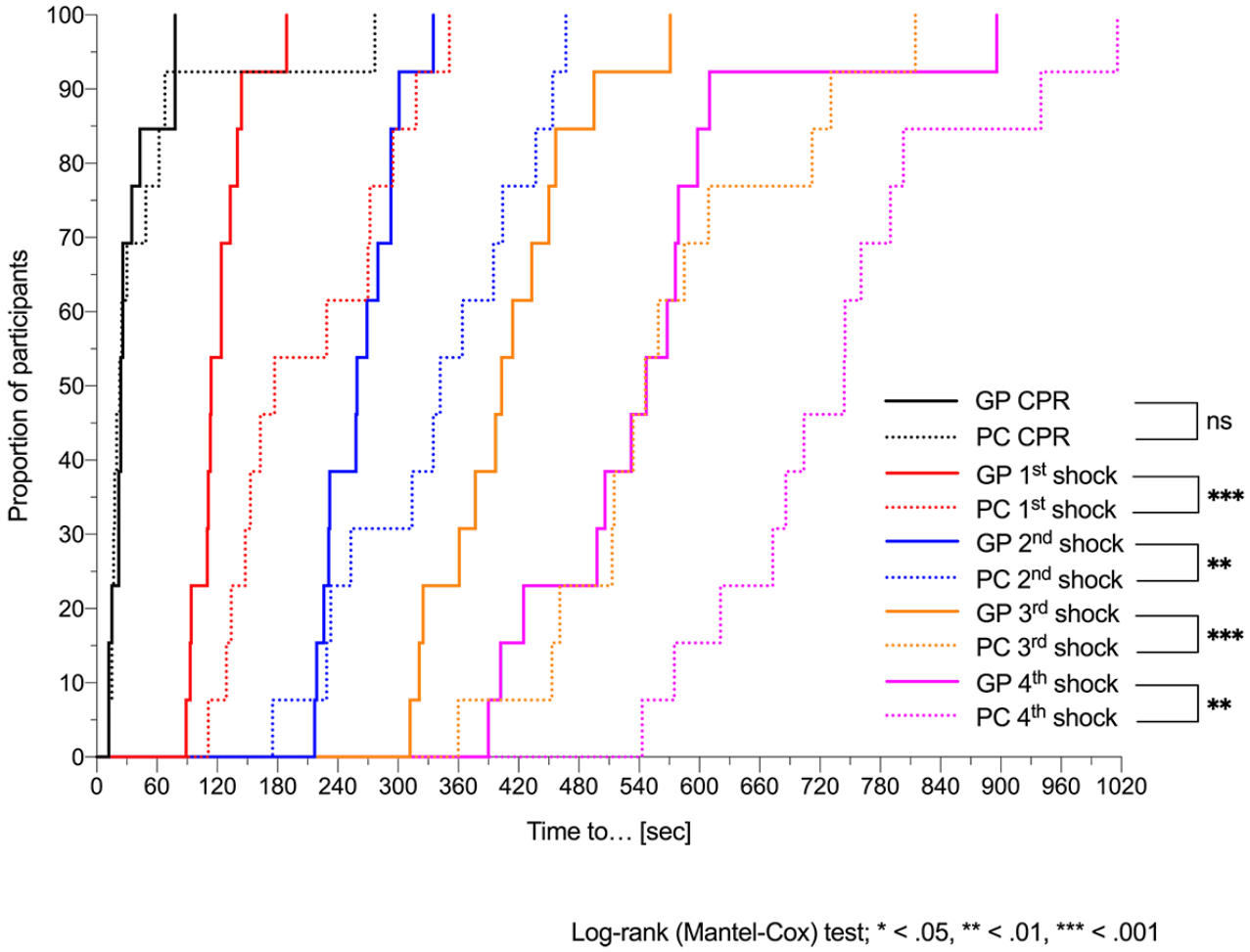


Table 2. Mean time to critical resuscitation endpoints.

Outcome	Mean time[a] for Guiding Pad app group (n=13), seconds		Mean time[a] for PALS[b] pocket cards group (n=13), seconds		Time difference[c], seconds	*P* value
	Mean (SD)	95% CI	Mean (SD)	95% CI		
Start chest compression	32.1 (22.1)	18.7-45.4	48.5 (71.1)	5.6-91.5	16.5	.91[d]
First defibrillation attempt	121.4 (26.7)	105.3-137.5	211.5 (81.2)	162.5-260.6	90.1	<.001
Intraosseous route	187.2 (45.4)	159.7-214.6	183.0 (71.3)	139.9-226.1	4.2	.86
Second defibrillation attempt	262.5 (36.7)	240.3-284.7	338.6 (93.8)	282.0-395.3	76.1	.01
Epinephrine	269.1 (74.8)	223.8-314.3	287.2 (82.9)	237.1-337.2	18.1	.56
Third defibrillation attempt	408.9 (74.1)	364.1-453.7	568.7 (124.4)	493.5-643.9	159.8	<.001
Amiodarone	455.5 (106.9)	391.0-520.1	598.2 (154.7)	504.7-691.7	142.7	.01
Fourth defibrillation attempt	548.2 (127.6)	471.1-625.4	738.5 (132.9)	658.2-818.9	190.3	<.001[d]

[a]The mean time in each allocation group refers to the delay in seconds from the end of the clinical statement given by the study investigator to each critical resuscitation endpoint.

[b]PALS: pediatric advanced life support.

[c]Time difference represents the absolute time difference between mean PALS pocket cards group and Guiding Pad app group outcomes.

[d]Mann-Whitney test.

Figure 8. Time to intraosseous (IO) route and drug delivery. Kaplan-Meier curves of time elapsed between the onset of simulated pulseless ventricular tachycardia (pVT) and the IO insertion, epinephrine, and amiodarone delivery for residents using the Guiding Pad (GP) app vs conventional pediatric advanced life support (PALS) pocket cards (PCs). Log-rank (Mantel-Cox) test comparing curves: χ^2=0.4 and *P*=.55 for the IO route; χ^2=0.6 and *P*=.44 for epinephrine; and χ^2=7.5 and *P*=.006 for amiodarone. ns: not significant.

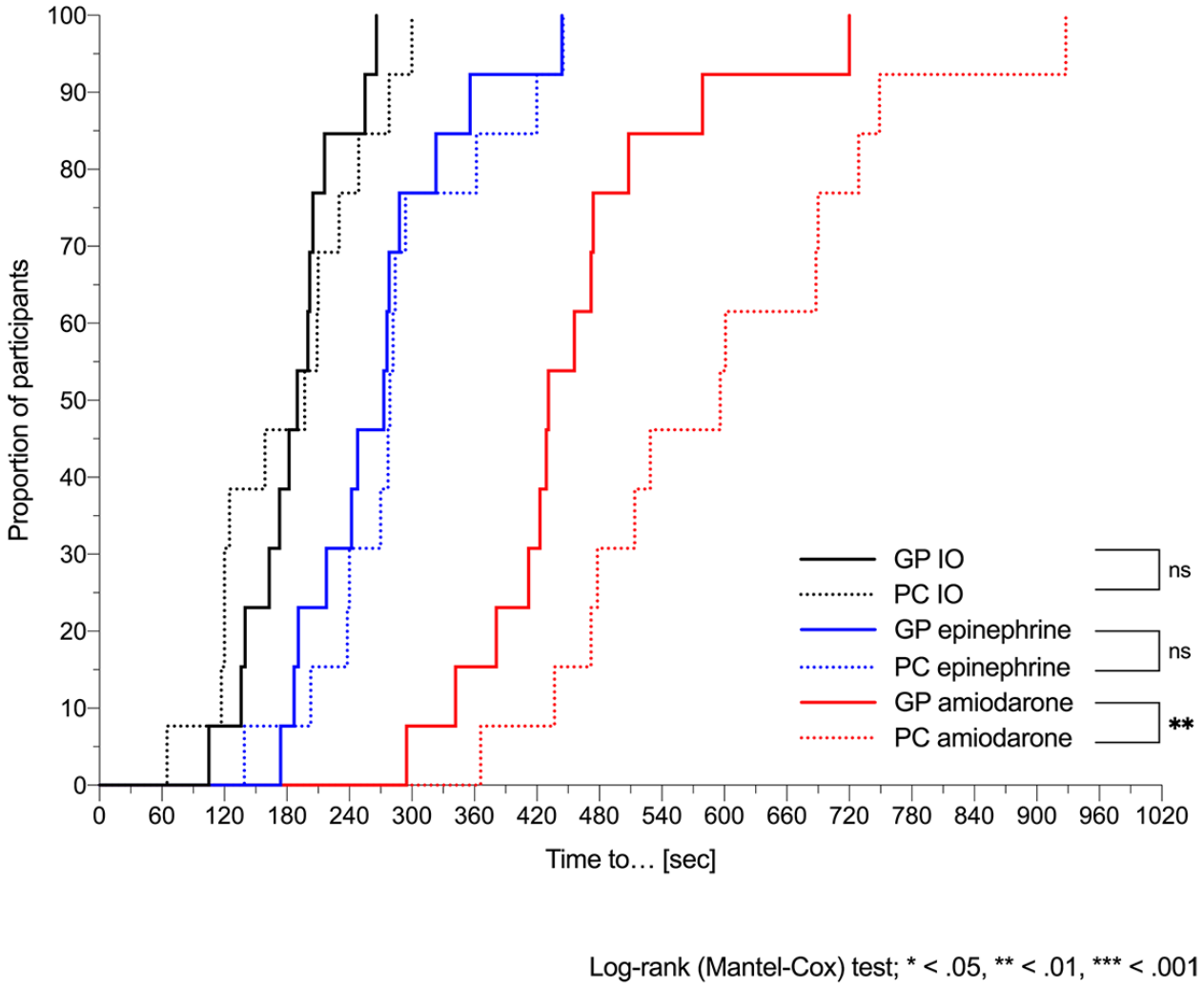


Figure 9. Association between time to defibrillation attempts and years of residency. Data are shown as a regression line (solid) with 95% CI (dashed lines). *P* values and r^2 values are based on simple linear regression analysis. White (Guiding Pad app) and grey (pediatric advanced life support [PALS] pocket cards) open circles denote each individual value.

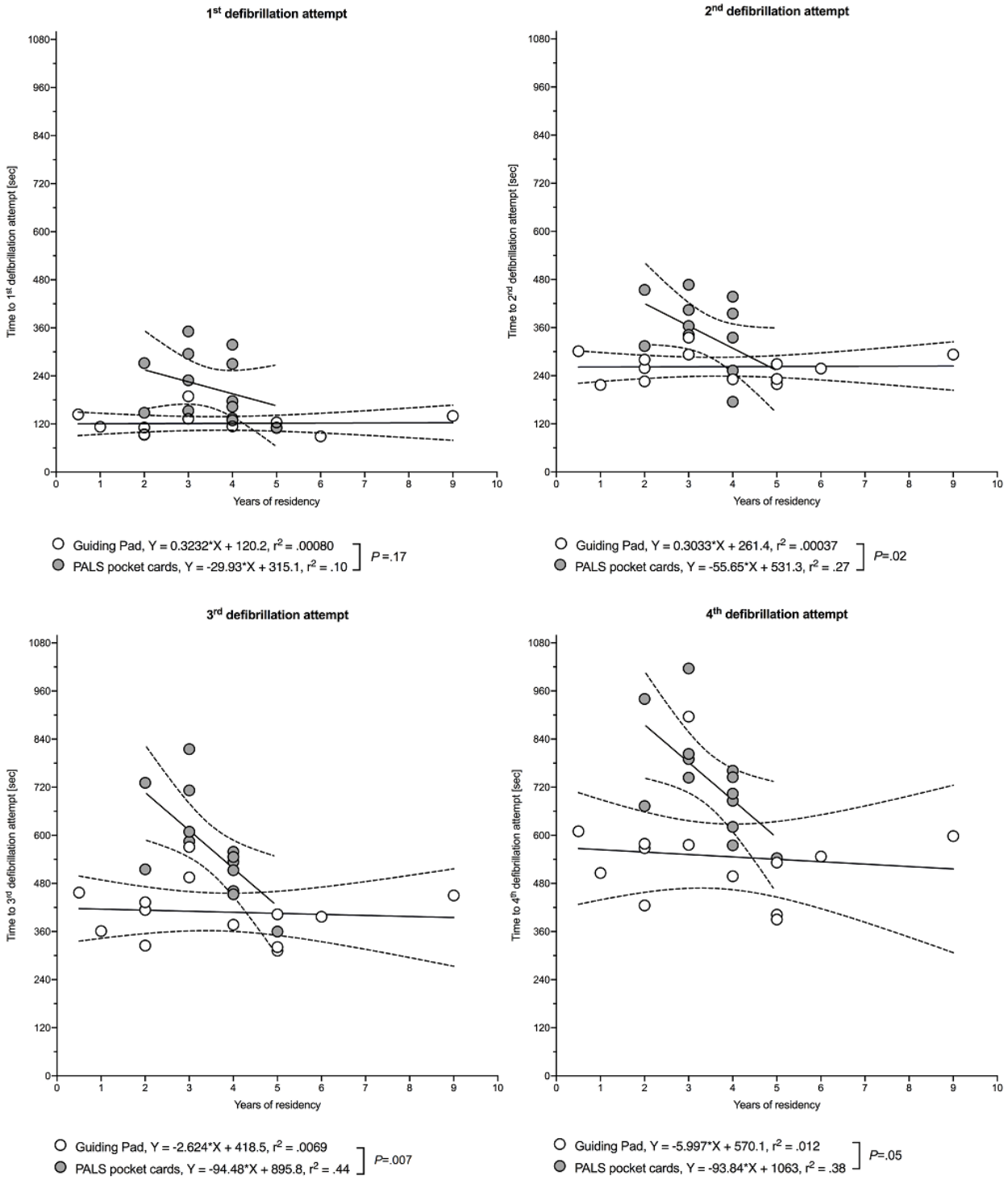
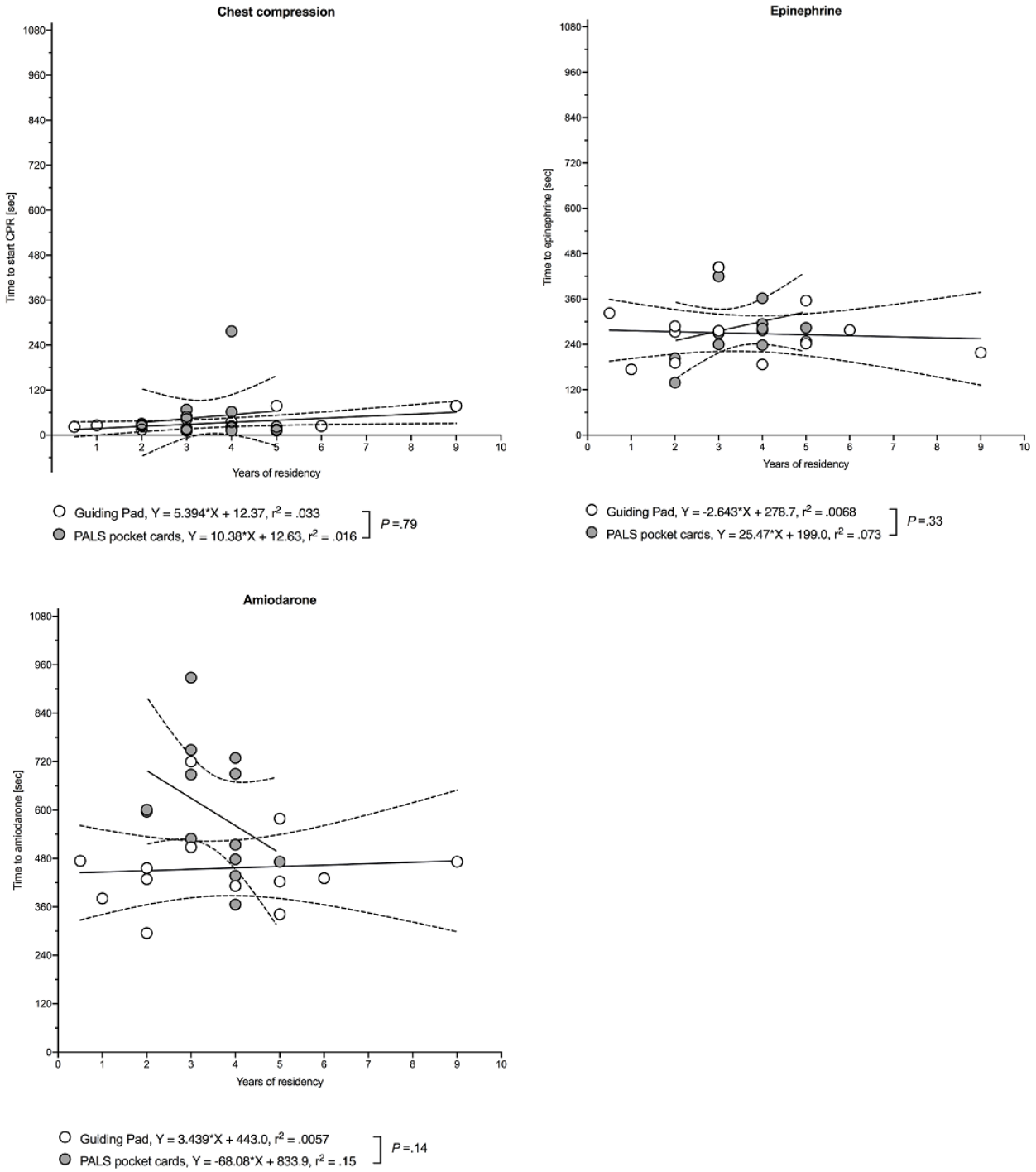


Figure 10. Association between time to chest compression or drug delivery and years of residency. Data are shown as a regression line (solid) with 95% CI (dashed lines). *P* values and r^2 values are based on simple linear regression analysis. White (Guiding Pad app) and grey (pediatric advanced life support [PALS] pocket cards) open circles denote each individual value. CPR: cardiopulmonary resuscitation.



Errors and Deviations From the AHA pVT Algorithm

Errors and deviations from the AHA pVT algorithm are summarized in Table 3 and in Table S1 in Multimedia Appendix 4. The entire pVT algorithm was followed correctly in a stepwise fashion until ROSC by 12 out of 13 (92%) residents in group A and only 3 out of 13 (23%) residents in group B (*P*=.001) (see Table 4). Importantly, the pVT rhythm was recognized correctly in 51 out of 52 opportunities (98%) by residents using the app, but in only 19 out of 52 opportunities (37%) of those using the pocket cards (*P*<.001).

Table 3. Errors and deviations from the American Heart Association (AHA) pulseless ventricular tachycardia (pVT) algorithm with respect to critical resuscitation endpoints.

Critical resuscitation endpoint	AHA recommended dose	Guiding Pad app (n=13)			PALS[a] pocket cards (n=13)		
		Dose, mean (SD)	95% CI	Dose range	Dose, mean (SD)	95% CI	Dose range
First defibrillation attempt (J/kg)	2.00	2.00 (0)	2.00-2.00	2.00-2.00	1.97 (0.76)	1.51-2.43	0.60-4.00
Second defibrillation attempt (J/kg)	4.00	3.85 (0.55)	3.51-4.18	2.00-4.00	3.62 (0.96)	3.04-4.20	1.00-4.00
Epinephrine 0.1 mg/mL (mL/kg)	0.10	0.10 (0)	0.10-0.10	0.10-0.10	0.09 (0.03)	0.08-0.11	0.001-0.10
Third defibrillation attempt (J/kg)	4.00	4.15 (0.55)	3.82-4.49	4.00-6.00	3.87 (1.57)	2.92-4.82	0.52-6.00
Amiodarone (mL/kg)	0.10	0.10 (0)	0.10-0.10	0.10-0.10	0.10 (0.01)	0.10-0.10	0.10-0.14
Fourth defibrillation attempt (J/kg)	4.00	4.31 (1.11)	3.64-4.98	4.00-8.00	5.45 (1.67)	4.44-6.46	4.00-8.00

[a]PALS: pediatric advanced life support.

Table 4. Errors and deviations from the American Heart Association (AHA) pulseless ventricular tachycardia (pVT) algorithm with respect to defibrillation and drug factors.

Defibrillation and drug factors	Guiding Pad app (n=13), n (%)	PALS[a] pocket cards (n=13), n (%)
Correct number of shocks (N=52)	51 (98)	51 (98)
Error in shock or drug doses (N=78)	1 (1)	11 (14)[b]
pVT rhythm recognition (N=52)	51 (98)	19 (37)[c]
Correct AHA sequence (n=13)	12 (92)	3 (23)[d]

[a]PALS: pediatric advanced life support.

[b]Difference between Guiding Pad app and PALS pocket cards groups: *P*=.005 (Fisher exact test).

[c]Difference between Guiding Pad app and PALS pocket cards groups: *P*<.001 (Fisher exact test).

[d]Difference between Guiding Pad app and PALS pocket cards groups: *P*=.001 (Fisher exact test).

Out of 52 opportunities, 1 error in the defibrillation dose (2%) was committed during the whole scenario in group A. This resident delivered a second asynchronous shock at half the recommended energy dose (2 J/kg instead of 4 J/kg). Owing to a discontinuous adherence to the app by switching alternatively with his own CPR experience, he also failed to comply with the algorithm and gave a mistimed 5 mg/kg dose of amiodarone 3 minutes after an unnecessary additional (2 J/kg) second defibrillation attempt. This compares to 8 out of 52 (15%) errors in defibrillation doses during the whole scenario in group B (*P*<.03): 3 at the first defibrillation attempt (doses ranged from 0.6 to 4 J/kg instead of 2 J/kg); 2 at the second attempt (1.0 to 2 J/kg instead of 4 J/kg); and 3 at the third attempt (0.52 to 2 J/kg instead of 4 to 10 J/kg) (see Table 3). Out of 13 residents, 2 in group B (15%) wrongly used synchronized shocks, either at the first, second, or third attempts. In group A, the mean energy dose of the first defibrillation attempt was strictly in accordance with the recommendations, whereas the second, third, and fourth defibrillation attempts deviated from the AHA recommendations by 0.15 J/kg (95% CI of discrepancy: –0.49 to 0.18, *P*=.34), 0.15 J/kg (95% CI of discrepancy: –0.18 to 0.49, *P*=.34), and 0.31 J/kg (95% CI of discrepancy: –0.36 to 0.98, *P*=.34), respectively. In group B, all four mean defibrillation attempts deviated from the AHA recommendations by 0.03 J/kg (95% CI of discrepancy: –0.49 to 0.43; *P*=.89), 0.38 J/kg (95% CI of discrepancy: –0.97 to 0.20; *P*=.17), 0.13 J/kg (95% CI of discrepancy: –1.08 to 0.82; *P*=.77), and 1.45 J/kg (95% CI of discrepancy: 0.44-2.46; *P*=.009), respectively.

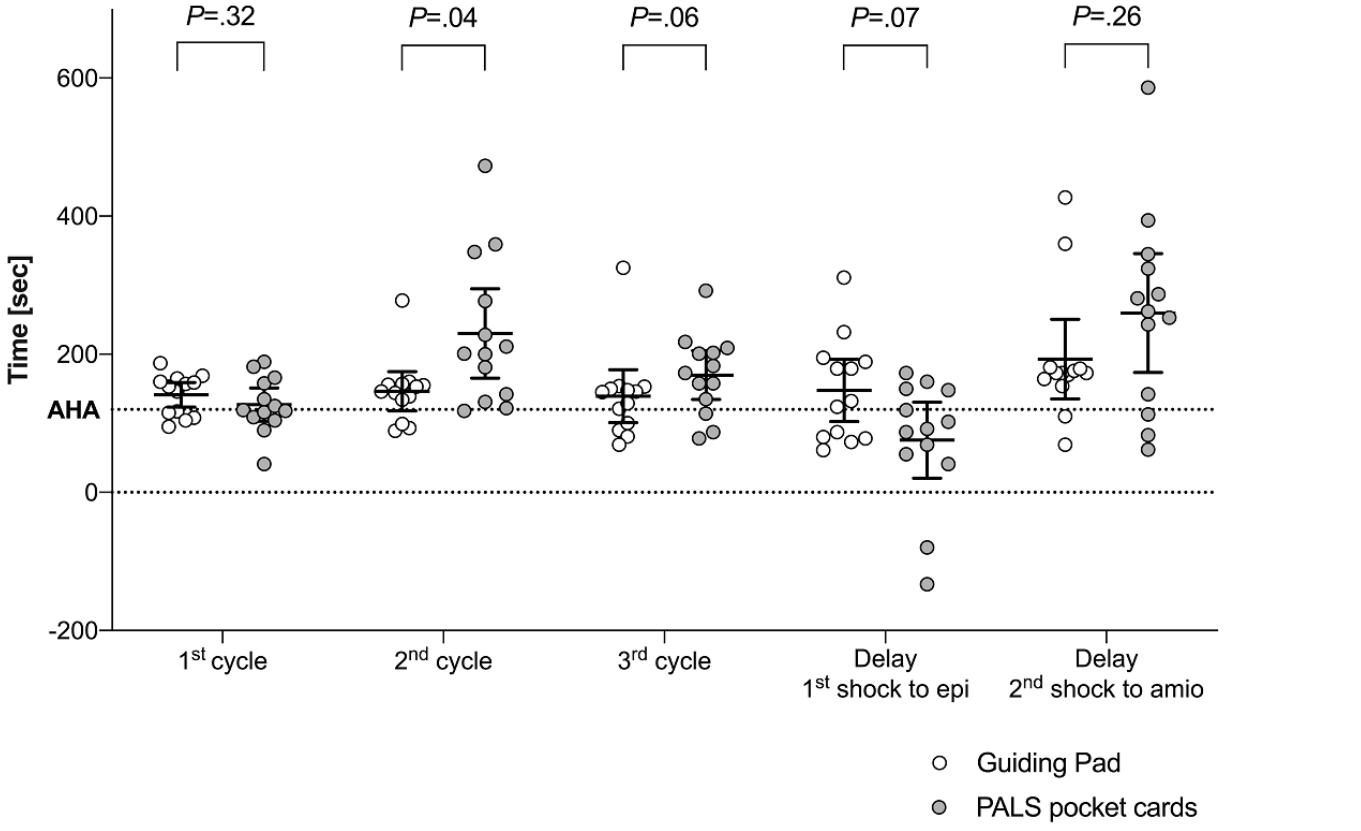
In group A, epinephrine drug doses were given according to AHA recommendations. However, in group B, epinephrine was delivered more than 2 minutes, on four occasions, either before the first (three times) or second (one time) shocks, and was once underdosed by 10 times the recommended dose. Regarding amiodarone, among card users, one resident wrongly ordered the drug before the first shock, another after the fourth shock, a third one at 1.4 times the recommended dose, and a resident even ordered a double dose before the fourth shock.

The hands-on time spent by cycles of chest compression between both groups is summarized in Figure 11. Using the Guiding Pad app, the mean time for the first, second, and third cycles of chest compression between each defibrillation attempt deviated from the AHA recommendations by 21.15 seconds (95% CI of discrepancy: 3.35-38.95; *P*=.02), 26.38 seconds (95% CI of discrepancy: –1.98 to 54.75; *P*=.07), and 19.30 seconds (95% CI of discrepancy: –18.88 to 57.49; *P*=.29), respectively. In group B, the mean time for the first, second, and third cycles of chest compression deviated from the AHA recommendations by 7.08 seconds (95% CI of discrepancy: –17.16 to 31.31; *P*=.54), 110.10 seconds (95% CI of discrepancy: 45.25-174.9; *P*=.003), and 49.85 seconds (95% CI of discrepancy: 14.58-85.11; *P*=.01), respectively. Mean delays between the first shock and epinephrine for the app and pocket card users

were 147.7 seconds (95% CI 102.6-192.7) and 75.6 seconds (95% CI 20.5-130.7), respectively (see Figure 11). Mean delays between the second shock and amiodarone for the app and pocket card users were 193.0 seconds (95% CI 135.2-250.8) and 259.6 seconds (95% CI 173.6-345.6), respectively (see Figure 11).

The questionnaire evaluating perceived stress and satisfaction scores was completed and returned by 100% of participants. Participants in groups A and B rated the overall perceived stress before the scenario with mean scores of 5.3 (95% CI 4.0-6.6) and 5.1 (95% CI 3.9-6.3), respectively (*P*=.78). During the scenario, the stress remained contained by the app users (mean score 4.8, 95% CI 3.4-6.2, *P*=.55), whereas it increased significantly for residents relying on the PALS pocket cards (mean score 6.8, 95% CI 5.9-7.8, *P*=.01) compared to app users (*P*=.01). Satisfaction tended to be greater for residents using the app (mean score 7.5, 95% CI 6.5-8.5) compared to those using pocket cards (mean score 5.9, 95% CI 4.4-7.4) (*P*=.07).

Figure 11. Time spent in seconds by cycles of chest compression and between defibrillation attempts and drug delivery during simulated pulseless ventricular tachycardia (pVT) scenarios. Solid horizontal lines denote mean and 95% CI. White (Guiding Pad app) and grey (pediatric advanced life support [PALS] pocket cards) open circles denote each individual value. The horizontal dashed line denotes the 120-sec American Heart Association (AHA) recommendation for a complete cycle. Delays between the first shock and epinephrine (epi) delivery, and between the second shock and amiodarone (amio) delivery, are expressed as the time to drug delivery minus the time to defibrillation attempt, by resident and by allocation group. A negative time point denotes a drug given before the expected defibrillation attempt.



Discussion

Principal Findings

In this randomized controlled trial, we report a reduced time to all defibrillation attempts and an improved adherence of approximately 70% to all CPR sequences of action outlined by the 2018 AHA pVT guidelines with the mobile app Guiding Pad compared with the PALS pocket reference cards among pediatric residents leading simulated CPR. Of note, this result was observed irrespective of residents' previous years of experience or prior CPR knowledge. Interindividual variance was also reduced with the app, suggesting a worthwhile benefit of its use by residents with various experience levels. To our knowledge, this is the first study to investigate the benefit of a mobile device app to improve the performance and adherence of pediatric residents to AHA resuscitation guidelines.

Standards of care acknowledge that a prompt defibrillation attempt is an important determinant of survival after cardiac arrest [27]. As outlined by Topijan et al, shorter duration of CPR is associated with higher rates of survival to discharge, supporting the concept of rapid recognition, prompt chest compression, and defibrillation as soon as possible [25]. During the first 15 minutes of CPR, survival and a favorable neurological outcome decrease linearly by 2.1% and 1.2% per minute, respectively [28]. Delays in initiating CPR have a detrimental effect on patient outcome, regardless of the quality of resuscitation [29]. Therefore, the AHA recommends that pulseless patients of any age should receive immediate CPR without delay starting with chest compressions followed by a defibrillation within 180 seconds of a shockable rhythm. In our study, approximately 80% of residents in both allocation groups started compressions within 60 seconds from the onset of pVT.

Importantly, mean time from initiation of chest compression to first shock was almost halved when using the app when compared to PALS pocket cards. Among residents using the app, 92% (12/13) defibrillated successfully in 180 seconds or less of pVT onset, whereas 46% (6/13) of PALS pocket card users failed to discharge the defibrillator within 180 seconds. This correlates well with the results of Hunt et al, who observed that despite the availability of AHA recommendations, 66% of pediatric residents failed to start compressions within 60 seconds from the onset of a simulated pVT, 33% never started compressions, only 54% successfully defibrillated within 180 seconds, and 7% never discharged the defibrillator [30]. A more recent study among first-year pediatric residents showed a median time of 50 seconds for the initiation of CPR and 282 seconds to first defibrillation [31]. Most alarmingly, the pVT rhythm in our trial was misidentified by almost 70% of residents holding the PALS reference cards in their hands. This could potentially negatively affect patient outcome as choosing the wrong electrical therapy, drugs, or algorithm in real life might impede the correct management of critically ill children and jeopardize their chance of survival.

Current AHA resuscitation guidelines emphasize 2 minutes of chest compressions between defibrillation attempts as optimal care for persistent pVT or ventricular fibrillation in children [32,33]. In this study, app users deviated less from the AHA, which reached statistical significance for the second and third cycles of chest compression. Moreover, following the first and second shocks and a 2-minute period of five cycles of CPR after each shock, antiarrhythmic drugs should be administered if the patient remains in cardiac arrest, with the aim of increasing defibrillation success with subsequent defibrillation attempts [32]. In this trial, both groups accurately administered epinephrine and amiodarone drug doses, with the exception of a 100-times underdosed epinephrine and a 1.4-times overdosed amiodarone dose in group B. On average, app users correctly respected a complete 2-minute cycle of chest compression-ventilation before administering epinephrine after the first shock. Conversely, and contrary to current AHA guidelines, pocket card users administered the drug too close to the first shock, possibly explaining the absence of a significant time difference to epinephrine administration between both groups, despite a significantly longer delay to deliver the first shock in group B. Due to further delays, amiodarone was delivered significantly later by more than 2 minutes among residents not using the app.

Prompt defibrillation is crucial for the termination of ventricular fibrillation or pVT in order to achieve ROSC [33]. The AHA 2018 guidelines recommend treating pVT or ventricular fibrillation in children with an initial dose of 2 J/kg [23]. For subsequent shocks, a dose of 4 J/kg is recommended, although higher energy levels may be considered up to an adult dose, if not exceeding 10 J/kg. In this trial, residents using the PALS pocket cards were more prone to deviate from defibrillation doses than those using the app. In a total of 52 defibrillation attempts, they deviated in 36% of cases. These deviations were reduced to 6% when using the app. It would be interesting in further studies to determine whether this would translate into fewer deviations in shock doses in real life.

While the app in this study offered better adherence to AHA resuscitation recommendations than conventional PALS pocket cards, we also found that it provided a relative advantage when compared to the Google Glass-based app dedicated to the same purpose [17]. The in-built small size of the screen was indeed a limiting factor reported by residents wearing the glasses, by hindering full display of algorithms. Usability issues were also observed with inopportune and time-consuming back-and-forth navigations throughout the algorithms. In this study, displaying the entire algorithm on the larger screen size of a tablet and paralleling stepwise patient-centered care guidance appeared to improve adherence to AHA guidelines and speed up skills, thus allowing residents to better manage simulated CPR. It would be interesting in further studies to assess this assumption with certified emergency physicians or paramedics in simulated and real-life in- or out-of-hospital environments. Given the evidence regarding the observed deviation from recommended resuscitation procedures, it might be also advisable to assess the educational impact of this app for the upstream training of rescuers' p-IHCA technical skills in further studies.

Limitations

Our study has some limitations. First, it was conducted during a resuscitation simulation–based scenario rather than tested in real-life situations. However, high-fidelity simulation is an essential method to teach resuscitation skills and technologies that cannot be practiced during real CPR, as the diversity among patients and their diseases makes such studies difficult to standardize in critical situations. The low occurrence of p-IHCA also limits the implementation of randomized trials in real life [34]. Moreover, standardizing the scenario and the environment helped to avoid effect modifiers by limiting the influence of undesired variables on the outcomes. Realism was achieved as reflected by the stress level experienced by participants, who considered the simulation to be as stressful as real CPR situations. Second, the 5-minute app training was dispensed just before the scenario. In real life, the interval between training and actual use would probably be months. However, training with the app months before the study would have unblinded participants to its purpose and could have created a preparation bias. Third, the sample size limited stratified analyses to estimate the impact of PALS certification on the outcomes, but a recent study observed that improved adherence to AHA recommendations was not directly associated with PALS-trained providers [7]. Finally, we acknowledge that our findings might not be generalizable to providers with extensive CPR experience, such as pediatric emergency physicians. As only residents were assessed in this trial, further studies would be valuable to assess this assumption.

Conclusions

A PALS-based mobile app designed for tablets to interactively support residents during pediatric CPR contributed to a shorter time to first and subsequent defibrillation attempts, fewer medication and defibrillation dose errors, as well as a better adherence to AHA recommendations, compared with the conventional PALS pocket reference cards. Taken together, our results suggest that residents are not accurately following AHA recommendations during pediatric CPR when only supported

by PALS pocket cards. A next step would be to determine, in real-life studies, whether this mobile app might benefit patients by improving the adherence and performance of residents to meet AHA resuscitation requirements in clinical practice.

Acknowledgments

We thank the residents for their contributions to the trial and Rosemary Sudan for providing editorial assistance. Geneva University Hospitals is the owner of the Guiding Pad app, which was not available at the time of submission on Google Play or the Apple App Store. This trial had financial support from the private foundation of Geneva University Hospitals (fund No. QS2-25). The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Authors' Contributions

JNS was responsible for the literature search and reading of articles, writing of the manuscript, preparation of figures and tables, and statistical analysis. JNS, LL, SM, and AC were responsible for data collection. FE was responsible for the development of the project software. JNS, LL, SM, AC, and FE were responsible for the concept and design of the study and critical review of manuscript content. All authors have contributed to, seen, and approved the final submitted version of the manuscript; had full access to all the data, including statistical reports and tables, in the study; and can take responsibility for the integrity of the data and the accuracy of the data analysis. The corresponding author (JNS) confirms that he had full access to the participants' data and endorsed the final responsibility for the submission. He further affirms that the manuscript is an honest, accurate, and transparent account of the study being reported, that no important aspects of the study have been omitted, and that any deviations from the study plan have been explained.

Conflicts of Interest

None declared.

This randomized study was not registered. The editor granted an exception as ICMJE does not require a registration if the purpose of the study is to examine the effect only on health care providers rather than patients. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Multimedia Appendix 1
CONSORT-eHEALTH V1.6.2.
[PDF File (Adobe PDF File), 174 KB - jmir_v22i5e17792_app1.pdf]

Multimedia Appendix 2
The pulseless ventricular tachycardia (pVT) resuscitation scenario.
[DOCX File , 13 KB - jmir_v22i5e17792_app2.docx]

Multimedia Appendix 3
Questionnaire for secondary outcomes, using 10-point Likert scales.
[DOCX File , 15 KB - jmir_v22i5e17792_app3.docx]

Multimedia Appendix 4
Interrater agreement on outcome analyses.
[DOCX File , 14 KB - jmir_v22i5e17792_app4.docx]

Multimedia Appendix 5
Bland and Altman analysis and intraclass correlation coefficient (ICC) on outcome analyses.
[DOCX File , 14 KB - jmir_v22i5e17792_app5.docx]

Multimedia Appendix 6
Bland and Altman analysis of pulseless ventricular tachycardia (pVT) algorithm review.
[DOCX File , 14520 KB - jmir_v22i5e17792_app6.docx]

Multimedia Appendix 7
Association between time to defibrillation attempts and number of prior cardiopulmonary resuscitation (CPR) attempts.
[DOCX File , 265 KB - jmir_v22i5e17792_app7.docx]

Multimedia Appendix 8
Association between time to chest compression or drug delivery and number of prior cardiopulmonary resuscitation (CPR) attempts.
[DOCX File , 191 KB - jmir_v22i5e17792_app8.docx]

References

1. Holmberg MJ, Ross CE, Fitzmaurice GM, Chan PS, Duval-Arnould J, Grossestreuer AV, American Heart Association's Get With The Guidelines–Resuscitation Investigators. Annual incidence of adult and pediatric in-hospital cardiac arrest in the United States. Circ Cardiovasc Qual Outcomes 2019 Jul 09;12(7):e005580. [Medline: 31545574]
2. Nadkarni VM, Larkin GL, Peberdy MA, Carey SM, Kaye W, Mancini ME, National Registry of Cardiopulmonary Resuscitation Investigators. First documented rhythm and clinical outcome from in-hospital cardiac arrest among children and adults. JAMA 2006 Jan 04;295(1):50-57. [doi: 10.1001/jama.295.1.50] [Medline: 16391216]
3. Benjamin EJ, Virani SS, Callaway CW, Chamberlain AM, Chang AR, Cheng S, American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics-2018 update: A report from the American Heart Association. Circulation 2018 Mar 20;137(12):e67-e492. [doi: 10.1161/CIR.0000000000000558] [Medline: 29386200]
4. López-Herce J, Del Castillo J, Matamoros M, Cañadas S, Rodriguez-Calvo A, Cecchetti C, Iberoamerican Pediatric Cardiac Arrest Study Network RIBEPCI. Factors associated with mortality in pediatric in-hospital cardiac arrest: A prospective multicenter multinational observational study. Intensive Care Med 2013 Feb;39(2):309-318. [doi: 10.1007/s00134-012-2709-7] [Medline: 23184036]
5. Cheng A, Nadkarni VM, Mancini MB, Hunt EA, Sinz EH, Merchant RM, American Heart Association Education Science Investigators; on behalf of the American Heart Association Education Science and Programs Committee, Council on Cardiopulmonary, Critical Care, Perioperative and Resuscitation; Council on Cardiovascular and Stroke Nursing; Council on Quality of Care and Outcomes Research. Resuscitation education science: Educational strategies to improve outcomes from cardiac arrest: A scientific statement from the American Heart Association. Circulation 2018 Aug 07;138(6):e82-e122. [doi: 10.1161/CIR.0000000000000583] [Medline: 29930020]
6. Hunt EA, Duval-Arnould JM, Nelson-McMillan KL, Bradshaw JH, Diener-West M, Perretta JS, et al. Pediatric resident resuscitation skills improve after "rapid cycle deliberate practice" training. Resuscitation 2014 Jul;85(7):945-951. [doi: 10.1016/j.resuscitation.2014.02.025] [Medline: 24607871]
7. Auerbach M, Brown L, Whitfill T, Baird J, Abulebda K, Bhatnagar A, et al. Adherence to pediatric cardiac arrest guidelines across a spectrum of fifty emergency departments: A prospective, in situ, simulation-based study. Acad Emerg Med 2018 Dec;25(12):1396-1408 [FREE Full text] [doi: 10.1111/acem.13564] [Medline: 30194902]
8. Nelson McMillan K, Rosen MA, Shilkofski NA, Bradshaw JH, Saliski M, Hunt EA. Cognitive aids do not prompt initiation of cardiopulmonary resuscitation in simulated pediatric cardiopulmonary arrests. Simul Healthc 2018 Feb;13(1):41-46. [doi: 10.1097/SIH.0000000000000297] [Medline: 29401183]
9. Meaney PA, Bobrow BJ, Mancini ME, Christenson J, de Caen AR, Bhanji F, CPR Quality Summit Investigators, the American Heart Association Emergency Cardiovascular Care Committee, and the Council on Cardiopulmonary, Critical Care, Perioperative and Resuscitation. Cardiopulmonary resuscitation quality: [Corrected] Improving cardiac resuscitation outcomes both inside and outside the hospital: A consensus statement from the American Heart Association. Circulation 2013 Jul 23;128(4):417-435. [doi: 10.1161/CIR.0b013e31829d8654] [Medline: 23801105]
10. Field LC, McEvoy MD, Smalley JC, Clark CA, McEvoy MB, Rieke H, et al. Use of an electronic decision support tool improves management of simulated in-hospital cardiac arrest. Resuscitation 2014 Jan;85(1):138-142 [FREE Full text] [doi: 10.1016/j.resuscitation.2013.09.013] [Medline: 24056391]
11. Low D, Clark N, Soar J, Padkin A, Stoneham A, Perkins GD, et al. A randomised control trial to determine if use of the iResus© application on a smart phone improves the performance of an advanced life support provider in a simulated medical emergency. Anaesthesia 2011 Apr;66(4):255-262 [FREE Full text] [doi: 10.1111/j.1365-2044.2011.06649.x] [Medline: 21401537]
12. Merchant RM, Abella BS, Abotsi EJ, Smith TM, Long JA, Trudeau ME, et al. Cell phone cardiopulmonary resuscitation: Audio instructions when needed by lay rescuers: A randomized, controlled trial. Ann Emerg Med 2010 Jun;55(6):538-543.e1. [doi: 10.1016/j.annemergmed.2010.01.020] [Medline: 20202719]
13. Semeraro F, Frisoli A, Loconsole C, Bannò F, Tammaro G, Imbriaco G, et al. Motion detection technology as a tool for cardiopulmonary resuscitation (CPR) quality training: A randomised crossover mannequin pilot study. Resuscitation 2013 Apr;84(4):501-507. [doi: 10.1016/j.resuscitation.2012.12.006] [Medline: 23238423]
14. Pérez Alonso N, Pardo Rios M, Juguera Rodriguez L, Vera Catalan T, Segura Melgarejo F, Lopez Ayuso B, et al. Randomised clinical simulation designed to evaluate the effect of telemedicine using Google Glass on cardiopulmonary resuscitation (CPR). Emerg Med J 2017 Nov;34(11):734-738. [doi: 10.1136/emermed-2016-205998] [Medline: 28768700]
15. Sondheim SE, Devlin J, Seward WH, Bernard AW, Feinn RS, Cone DC. Recording out-of-hospital cardiac arrest treatment via a mobile smartphone application: A feasibility simulation study. Prehosp Emerg Care 2019;23(2):284-289. [doi: 10.1080/10903127.2018.1490838] [Medline: 30130413]

16. Drummond D, Arnaud C, Guedj R, Duguet A, de Suremain N, Petit A. Google Glass for residents dealing with pediatric cardiopulmonary arrest: A randomized, controlled, simulation-based study. Pediatr Crit Care Med 2017 Feb;18(2):120-127. [doi: 10.1097/PCC.0000000000000977] [Medline: 28165347]
17. Siebert JN, Ehrler F, Gervaix A, Haddad K, Lacroix L, Schrurs P, et al. Adherence to AHA guidelines when adapted for augmented reality glasses for assisted pediatric cardiopulmonary resuscitation: A randomized controlled trial. J Med Internet Res 2017 May 29;19(5):e183 [FREE Full text] [doi: 10.2196/jmir.7379] [Medline: 28554878]
18. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: Improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
19. Cheng A, Kessler D, Mackinnon R, Chang TP, Nadkarni VM, Hunt EA, International Network for Simulation-based Pediatric Innovation, Research, and Education (INSPIRE) Reporting Guidelines Investigators. Reporting guidelines for health care simulation research: Extensions to the CONSORT and STROBE statements. Simul Healthc 2016 Aug;11(4):238-248. [doi: 10.1097/SIH.0000000000000150] [Medline: 27465839]
20. Sealed Envelope. URL: https://www.sealedenvelope.com/ [accessed 2020-05-11]
21. Samson RA, Nadkarni VM, Meaney PA, Carey SM, Berg MD, Berg RA, American Heart Association National Registry of CPR Investigators. Outcomes of in-hospital ventricular fibrillation in children. N Engl J Med 2006 Jun 01;354(22):2328-2339. [doi: 10.1056/NEJMoa052917] [Medline: 16738269]
22. Siebert JN, Ehrler F, Combescure C, Lovis C, Haddad K, Hugon F, PedAMINES Trial Group. A mobile device application to reduce medication errors and time to drug delivery during simulated paediatric cardiopulmonary resuscitation: A multicentre, randomised, controlled, crossover trial. Lancet Child Adolesc Health 2019 May;3(5):303-311. [doi: 10.1016/S2352-4642(19)30003-3] [Medline: 30797722]
23. Duff JP, Topjian A, Berg MD, Chan M, Haskell SE, Joyner BL, et al. 2018 American Heart Association focused update on pediatric advanced life support: An update to the American Heart Association guidelines for cardiopulmonary resuscitation and emergency cardiovascular care. Circulation 2018 Dec 04;138(23):e731-e739. [doi: 10.1161/CIR.0000000000000612] [Medline: 30571264]
24. Bircher NG, Chan PS, Xu Y, American Heart Association's Get With The Guidelines–Resuscitation Investigators. Delays in cardiopulmonary resuscitation, defibrillation, and epinephrine administration all decrease survival in in-hospital cardiac arrest. Anesthesiology 2019 Mar;130(3):414-422. [doi: 10.1097/ALN.0000000000002563] [Medline: 30707123]
25. Topjian A. Shorter time to defibrillation in pediatric CPR: Children are not small adults, but shock them like they are. JAMA Netw Open 2018 Sep 07;1(5):e182653 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.2653] [Medline: 30646162]
26. Dilley SJ, Weiland TJ, O'Brien R, Cunningham NJ, Van Dijk JE, Mahoney RM, et al. Use of a checklist during observation of a simulated cardiac arrest scenario does not improve time to CPR and defibrillation over observation alone for subsequent scenarios. Teach Learn Med 2015;27(1):71-79. [doi: 10.1080/10401334.2014.979182] [Medline: 25584474]
27. Chan PS, Krumholz HM, Nichol G, Nallamothu BK, American Heart Association National Registry of Cardiopulmonary Resuscitation Investigators. Delayed time to defibrillation after in-hospital cardiac arrest. N Engl J Med 2008 Jan 03;358(1):9-17. [doi: 10.1056/NEJMoa0706467] [Medline: 18172170]
28. Matos RI, Watson RS, Nadkarni VM, Huang H, Berg RA, Meaney PA, American Heart Association's Get With The Guidelines–Resuscitation (Formerly the National Registry of Cardiopulmonary Resuscitation) Investigators. Duration of cardiopulmonary resuscitation and illness category impact survival and neurologic outcomes for in-hospital pediatric cardiac arrests. Circulation 2013 Jan 29;127(4):442-451. [doi: 10.1161/CIRCULATIONAHA.112.125625] [Medline: 23339874]
29. Ross JC, Trainor JL, Eppich WJ, Adler MD. Impact of simulation training on time to initiation of cardiopulmonary resuscitation for first-year pediatrics residents. J Grad Med Educ 2013 Dec;5(4):613-619 [FREE Full text] [doi: 10.4300/JGME-D-12-00343.1] [Medline: 24455010]
30. Hunt EA, Vera K, Diener-West M, Haggerty JA, Nelson KL, Shaffner DH, et al. Delays and errors in cardiopulmonary resuscitation and defibrillation by pediatric residents during simulated cardiopulmonary arrests. Resuscitation 2009 Jul;80(7):819-825. [doi: 10.1016/j.resuscitation.2009.03.020] [Medline: 19423210]
31. Jeffers J, Eppich W, Trainor J, Mobley B, Adler M. Development and evaluation of a learning intervention targeting first-year resident defibrillation skills. Pediatr Emerg Care 2016 Apr;32(4):210-216. [doi: 10.1097/PEC.0000000000000765] [Medline: 26990849]
32. de Caen AR, Berg MD, Chameides L, Gooden CK, Hickey RW, Scott HF, et al. Part 12: Pediatric Advanced Life Support: 2015 American Heart Association guidelines update for cardiopulmonary resuscitation and emergency cardiovascular care. Circulation 2015 Nov 03;132(18 Suppl 2):S526-S542 [FREE Full text] [doi: 10.1161/CIR.0000000000000266] [Medline: 26473000]
33. Kleinman ME, Chameides L, Schexnayder SM, Samson RA, Hazinski MF, Atkins DL, et al. Part 14: Pediatric Advanced Life Support: 2010 American Heart Association guidelines for cardiopulmonary resuscitation and emergency cardiovascular care. Circulation 2010 Nov 02;122(18 Suppl 3):S876-S908. [doi: 10.1161/CIRCULATIONAHA.110.971101] [Medline: 20956230]

34. Cheng A, Auerbach M, Hunt EA, Chang TP, Pusic M, Nadkarni V, et al. Designing and conducting simulation-based research. Pediatrics 2014 Jun;133(6):1091-1101. [doi: 10.1542/peds.2013-3267] [Medline: 24819576]

Abbreviations

AHA: American Heart Association
BLS: basic life support
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online TeleHealth
CPR: cardiopulmonary resuscitation
PALS: pediatric advanced life support
PED: pediatric emergency department
p-IHCA: pediatric in-hospital cardiac arrest
p-OHCA: pediatric out-of-hospital cardiac arrest
pVT: pulseless ventricular tachycardia
ROSC: return of spontaneous circulation

Edited by G Eysenbach; submitted 13.01.20; peer-reviewed by A Benis, B Munzer, KA Nguyen; comments to author 18.03.20; revised version received 22.03.20; accepted 23.03.20; published 27.05.20.

Please cite as:
Siebert JN, Lacroix L, Cantais A, Manzano S, Ehrler F
The Impact of a Tablet App on Adherence to American Heart Association Guidelines During Simulated Pediatric Cardiopulmonary Resuscitation: Randomized Controlled Trial
J Med Internet Res 2020;22(5):e17792
URL: http://www.jmir.org/2020/5/e17792/
doi:10.2196/17792
PMID:32292179

Original Paper

Use of Smartphones to Detect Diabetic Retinopathy: Scoping Review and Meta-Analysis of Diagnostic Test Accuracy Studies

Choon Han Tan[1]; Bhone Myint Kyaw[2], MBBS, MSc, PhD; Helen Smith[3], MSc, DM, MRCGP, FFPHM; Colin S Tan[1,4], MBBS, FRCSEd, MMed; Lorainne Tudor Car[3,5], MD, MSc, PhD

[1]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

[2]Centre for Population Health Sciences, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

[3]Family Medicine and Primary Care, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

[4]Department of Ophthalmology, Tan Tock Seng Hospital, Singapore, Singapore

[5]Department of Primary Care and Public Health, School of Public Health, Imperial College London, London, United Kingdom

Corresponding Author:
Lorainne Tudor Car, MD, MSc, PhD
Family Medicine and Primary Care, Lee Kong Chian School of Medicine, Nanyang Technological University
11 Mandalay Road
Singapore
Singapore
Phone: 65 6904 1258
Email: lorainne.tudor.car@ntu.edu.sg

Abstract

Background: Diabetic retinopathy (DR), a common complication of diabetes mellitus, is the leading cause of impaired vision in adults worldwide. Smartphone ophthalmoscopy involves using a smartphone camera for digital retinal imaging. Utilizing smartphones to detect DR is potentially more affordable, accessible, and easier to use than conventional methods.

Objective: This study aimed to determine the diagnostic accuracy of various smartphone ophthalmoscopy approaches for detecting DR in diabetic patients.

Methods: We performed an electronic search on the Medical Literature Analysis and Retrieval System Online (MEDLINE), EMBASE, and Cochrane Library for literature published from January 2000 to November 2018. We included studies involving diabetic patients, which compared the diagnostic accuracy of smartphone ophthalmoscopy for detecting DR to an accurate or commonly employed reference standard, such as indirect ophthalmoscopy, slit-lamp biomicroscopy, and tabletop fundus photography. Two reviewers independently screened studies against the inclusion criteria, extracted data, and assessed the quality of included studies using the Quality Assessment of Diagnostic Accuracy Studies–2 tool, with disagreements resolved via consensus. Sensitivity and specificity were pooled using the random effects model. A summary receiver operating characteristic (SROC) curve was constructed. This review is reported in line with the Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies guidelines.

Results: In all, nine studies involving 1430 participants were included. Most studies were of high quality, except one study with limited applicability because of its reference standard. The pooled sensitivity and specificity for detecting any DR was 87% (95% CI 74%-94%) and 94% (95% CI 81%-98%); mild nonproliferative DR (NPDR) was 39% (95% CI 10%-79%) and 95% (95% CI 91%-98%); moderate NPDR was 71% (95% CI 57%-81%) and 95% (95% CI 88%-98%); severe NPDR was 80% (95% CI 49%-94%) and 97% (95% CI 88%-99%); proliferative DR (PDR) was 92% (95% CI 79%-97%) and 99% (95% CI 96%-99%); diabetic macular edema was 79% (95% CI 63%-89%) and 93% (95% CI 82%-97%); and referral-warranted DR was 91% (95% CI 86%-94%) and 89% (95% CI 56%-98%). The area under SROC curve ranged from 0.879 to 0.979. The diagnostic odds ratio ranged from 11.3 to 1225.

Conclusions: We found heterogeneous evidence showing that smartphone ophthalmoscopy performs well in detecting DR. The diagnostic accuracy for PDR was highest. Future studies should standardize reference criteria and classification criteria and evaluate other available forms of smartphone ophthalmoscopy in primary care settings.

(J Med Internet Res 2020;22(5):e16658) doi:10.2196/16658

KEYWORDS

diabetic retinopathy; smartphone; mobile phone; ophthalmoscopy; artificial intelligence; telemedicine

Introduction

Diabetic retinopathy (DR) is the leading cause of impaired vision worldwide [1]. One in three patients with diabetes mellitus (DM) have DR [2]. DR includes proliferative DR (PDR) and various levels of nonproliferative DR (NPDR). PDR, characterized by retinal neovascularization at the disc and elsewhere, displays signs of angiogenesis in response to retinal tissue hypoxia. Neovascularization potentially leads to preretinal and vitreous hemorrhage, resulting in visual loss and, eventually, tractional retinal detachment. It may also cause iris neovascularization with resultant increase in intraocular pressure, eventually leading to neovascular glaucoma [3]. Typical clinical features of NPDR include the following: (1) microaneurysms and intraretinal hemorrhages from weak capillary walls; (2) hard exudates from vascular protein leakage; and (3) cotton wool spots, caused by ischemic infarcts leading to fluid accumulation. Diabetic macular edema (DME), caused by the thickening of and fluid accumulation in the retina, can occur at any stage of DR [4].

Diabetic eye disease is treatable. Treatments include vascular endothelial growth factor inhibitors, panretinal or focal photocoagulation, and vitrectomy [5]. Strict glycemic and blood pressure control can also delay the development of DR or reduce DR severity [6]. Treatments available are more effective at halting or slowing visual loss than reversing visual impairment. Yet, most patients remain asymptomatic until the advanced stages of DR. Therefore, early detection of DR before irreversible loss of visual acuity is crucial to ensure better patient outcomes [7].

The gold standard diagnostic test for DR is the Early Treatment Diabetic Retinopathy Study (ETDRS) 7-field stereoscopic color fundus photography or fluorescein angiography [8]. However, fundus cameras are nonportable, expensive, and operator dependent, often requiring patients to sit upright [9,10]. Moreover, fluorescein angiography is invasive, costly, and associated with prominent side effects [11]. Thus, they are impractical for screening in primary care or mobile settings. Other accurate [12] and frequently employed DR identification approaches include the following: (1) ophthalmoscopy; (2) slit-lamp biomicroscopy; and (3) other forms of fundus photography [13]. Optical coherence tomography is an emerging technology that reliably identifies DME by quantifying retinal thickness [14], but it is expensive and bulky and it cannot accurately grade DR severity.

Smartphone ophthalmoscopy, the use of a smartphone's in-built camera for retinal imaging, could be a valuable method for detecting DR because of its affordability, portability, and ease of use compared with traditional approaches [15]. Various health care workers could potentially operate a smartphone-based retinal imaging device, without limiting this procedure to highly specialized staff. Images acquired by smartphones can be easily shared with and graded remotely by ophthalmologists or other trained graders via telemedicine. These benefits are particularly important in resource-constrained health care settings, such as rural areas in developing countries lacking medical equipment and trained health care professionals [16]. Several literature reviews [17-19] have discussed smartphone retinal imaging technology and underscored the huge potential of smartphone ophthalmoscopy for detecting DR. Given the potential of this novel approach, we performed a scoping review to systematically collate and assess evidence regarding the accuracy of smartphone ophthalmoscopy for DR identification.

Methods

Reporting Guidelines

This scoping review was reported in line with the Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies (PRISMA-DTA) guidelines [20] and conducted according to the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy [21]. We adopted a scoping review approach [22,23] because of a broad set of inclusion criteria. The protocol for this review was published in *BMJ Open* [24]. We were unable to register this protocol with PROSPERO as it does not include scoping reviews.

Search Strategy

We performed a librarian-assisted search on the Medical Literature Analysis and Retrieval System Online (MEDLINE) (Ovid), EMBASE (Ovid), and the Cochrane Library for papers published from January 2000 to November 2018. Articles published before 2000 were excluded because before that smartphone technology was limited. We used both medical subject headings (MeSH) and keywords relating to DR (eg, "diabetic retinopathy," "macular edema," and "diabetic maculopathy") and to smartphones (eg, "mobile health," "mobile phones," and "applications") or AI (eg, "artificial intelligence" and "machine learning"; Multimedia Appendix 1). We also explored the bibliography of both primary articles and reviews to identify potentially eligible studies missed by the electronic search.

Study Selection

The inclusion criteria were as follows: (1) studies evaluating the diagnostic test accuracy of smartphone ophthalmoscopy for detecting DR in patients with type 1 or 2 DM; (2) studies utilizing a smartphone's in-built camera for retinal imaging, including the use of any attachments externally fitted to the smartphone; (3) studies comparing smartphone ophthalmoscopy with any acceptable and commonly employed reference standard, such as fundus photography, indirect ophthalmoscopy, slit-lamp biomicroscopy, or fluorescein angiography; (4) studies employing any kind of health care professional to acquire the smartphone images. Language was not an exclusion criterion.

Examples of eligible smartphone ophthalmoscopy techniques include the following:

- Direct ophthalmoscopy: An adaptor is externally attached to a smartphone's camera. These adaptors usually contain polarizers that reduce artifacts from corneal reflections. The arrangement of polarizers, beam-splitters, and lenses produces an annular illumination pattern.
- Indirect ophthalmoscopy: This simpler, monocular design involves a single lens (eg, 20 D condenser) placed between the smartphone camera and eye. It can be mounted on the phone via hardware or manually held in position.

Covidence software (Veritas Health Innovation, Melbourne, Australia) was used to remove duplicated studies [25]. After a pilot screening of 20 citations to calibrate the judging criteria, two reviewers independently screened all articles retrieved from the search strategy by title and abstract, using Covidence. Subsequently, we screened the full text of the remaining articles and performed data extraction using a prepiloted form. Any disagreements were resolved through consensus.

Data Collection

A data extraction form (Multimedia Appendix 2) was created and piloted to record the following data from each study: (1) study author and date published; (2) sample size; (3) participant characteristics (eg, age, duration and type of DM); (4) information regarding imaging techniques (eg, details about smartphones and adaptors used, image resolution); (5) health care professional performing smartphone ophthalmoscopy; (6) reference standard used; and (7) test results (eg, true positives [TP], false positives [FP], true negatives [TN], and false negatives [FN]). Corresponding authors were contacted for additional details or missing data required to construct a 2×2 table. Two reviewers independently extracted study data using a data extraction template created in Microsoft Excel, with disagreements resolved via consensus.

Quality Assessment

The Quality Assessment of Diagnostic Accuracy Studies tool, QUADAS-2, consisting of descriptions and signaling questions, was used to assess the risk of bias and applicability of all included studies in four domains pertaining to (1) patient selection, (2) index test, (3) reference standard, and (4) flow of participants through the study and timing between the index test and reference standard [26]. Two reviewers independently assessed study quality, and disagreements were resolved via discussion until a consensus was reached.

Statistical Analysis

We constructed 2×2 tables based on data from each study. The sensitivity, specificity, positive likelihood ratio (LR+), negative likelihood ratio (LR−), diagnostic odds ratio (DOR), and area under summary receiver operating characteristic (SROC) curve were calculated using a random effects model because of the high expected heterogeneity [27]. We constructed SROC using the bivariate model where possible. Being both a hierarchical and random effects model, the bivariate model is preferred to the Moses-Littenberg SROC curve—the former method accounts for between-study heterogeneity. For SROC curves employing the bivariate model, elliptical 95% confidence regions were obtained by joining the individual confidence regions for logit sensitivity and logit specificity via parametric representations [28,29].

Heterogeneity was evaluated using chi-square (χ^2) and I^2 values of likelihood ratio tests (LRT) or DOR, with I^2<25%, 25–75%, and >75% representing low, moderate, and high degree of inconsistency, respectively. Threshold effect was measured using the Spearman correlation coefficient ρ between logits of sensitivity and specificity, with ρ closer to −1 indicating higher threshold effect and better fit of the SROC curve. If information regarding a condition's prevalence was available from the literature, we calculated the posttest probability using the Fagan nomogram. A P<.05 was considered statistically significant. All analyses were performed using Review Manager version 5.3 from the Cochrane Collaboration [30], METANDI and MIDAS commands in Stata 15.1 (StataCorp, College Station, Texas), Meta-Disc version 1.4 (Ramón y Cajal Hospital, Madrid, Spain) [31], and *mada* package in R.

Results

Study Selection and Study Characteristics

Our search strategy yielded 1571 unique records. Of those records, the full text for 41 articles was assessed, and nine studies [32-40] met the inclusion criteria (Figure 1). Two [32,33] of the included studies were conference abstracts. A total of 1430 diabetic patients (at least 2743 eyes) were recruited among these studies.

All studies reported smartphone fundoscopy techniques involving mydriatic, color, and nonstereoscopic imaging (Tables 1 and 2, Multimedia Appendix 3). A total of four studies originated from India, three studies from the United States, and one from Italy. All studies which reported data on gender recruited both males and females. These fundus photographs were graded by ophthalmologists, retinal specialists, or artificial intelligence (AI). In all, seven studies utilized direct ophthalmoscopy to acquire smartphone fundus images, while two studies [38,40] used indirect ophthalmoscopy.

A total of five studies [32,33,37,39,40] employed slit-lamp biomicroscopy as the reference standard, of which two studies complemented the examination with dilated indirect ophthalmoscopy [39,40]. In all, two studies [35,38] utilized 7-field mydriatic fundus photography using a tabletop fundus camera; one study [34] used *traditional in-clinic diagnosis* including a dilated eye examination; and one study [36] utilized ophthalmologists' grading of the same smartphone-acquired images as the reference standard. Overall, four studies [32,36,37,40] utilized the International Clinical DR Disease Severity Scale to grade DR; three studies [34,35,38] employed the Airlie House or modified ETDRS criteria; and one study [39] used the United Kingdom's National Health Service (NHS) guidelines. Referral-warranted DR (RWDR) was defined as moderate NPDR or worse or DME; vision-threatening DR (VTDR) as severe NPDR, PDR, or DME; and sight-threatening DR (STDR) as PDR or DME. Health professionals performing smartphone ophthalmoscopy included medical students, interns or assistants, retinal specialists, ophthalmologists, and ophthalmic photographers. Most studies reported no funding

sources or conflicting interests, while such information was unavailable for two studies [32,33]. In all, two studies [34,40] received funding, one of which disclosed multiple authors holding positions in DigiSight Technologies, Inc.

Figure 1. Flowchart depicting the identification of relevant studies.

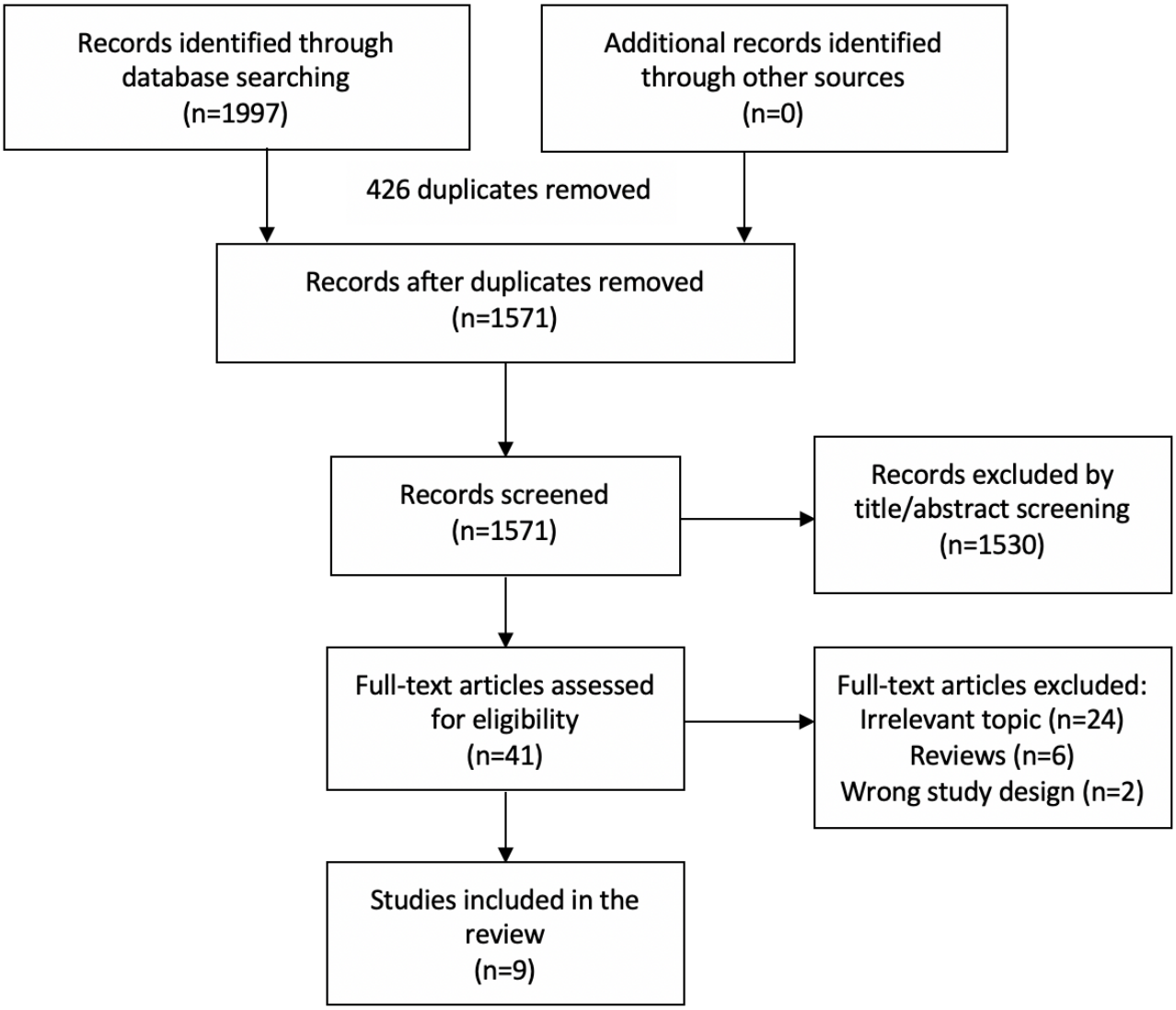


Table 1. Characteristics of included studies.

Study author, year	Country, setting	Sample size (patients/eyes)	Age (years), mean (SD)	Diabetes duration (years)		Diabetic retinopathy severity scale	Reference standard
				Mean (SD)	Range		
Bhat, 2016 [32]	N/A[a]	80/N/A	N/A	N/A	N/A	ICDR[b] severity scale; no referral defined as no or mild signs of DR[c].	Slit-lamp exam
Kim, 2017 [33]	United States, Retina Clinic	72/144	N/A	N/A	N/A	Referable DR defined as moderate NPDR[d] or worse, or DME[e].	Slit-lamp biomicroscopy
Kim, 2018 [34]	United States, Retina Clinic	71/142	56.7 (16.9)	N/A	N/A	Airlie House ETDRS[f] criteria; RWDR[g] defined as moderate NPDR or worse, or DME.	Gold standard dilated eye examination, with optical coherence tomography for DME
Rajalakshmi, 2015 [35]	India, Tertiary care diabetes hospital	301/602	53.5 (9.6)	12.5 (7.3)	N/A	Modified ETDRS criteria; STDR[h] defined as PDR[i] or DME	Mydriatic 7-standard field digital retinal photography
Rajalakshmi, 2018 [36]	India, Tertiary care diabetes hospital	301/602	N/A	N/A	N/A	ICDR severity scale; STDR defined as severe NPDR, PDR, or DME; RDR[j] defined as moderate NPDR or worse, or DME.	Remidio Fundus On Phone images graded by ophthalmologists
Russo, 2015 [37]	Italy, Diabetic center	120/240	58.8 (16.4)	11.6 (9.7)	N/A	ICDR severity scale; ETDRS criteria for DME; RWDR defined as moderate NPDR or worse, regardless of DME status.	Dilated slit-lamp biomicroscopy by a retinal specialist
Ryan, 2015 [38]	India, Ophthalmology clinic of a tertiary diabetes care center	300/600	48.0 (11.0)	N/A	0.1-37.2 years	Modified ETDRS criteria; VTDR[k] defined as severe NPDR or worse, or DME.	Mydriatic 7-field fundus photography by trained optometrists
Sengupta, 2018 [39]	India, Aravind Eye Hospital	135/233	54.1 (8.3)	10.7 (5.1)	N/A	National Health Service guidelines; VTDR defined as R2-level or worse (severe NPDR, PDR), or DME.	Dilated slit-lamp biomicroscopy (+90 D lens) and indirect ophthalmoscopy by retinal specialists
Toy, 2016 [40]	United States, Health care safety-net ophthalmology clinic	50/100	60.5 (10.6)	11.9 (8.4)	N/A	ICDR severity scale; RWDR defined as moderate NPDR or worse, or ungradable images.	Slit-lamp exam + dilated ophthalmoscopy by technicians

[a]N/A: not available.

[b]ICDR: International Clinical Diabetic Retinopathy.

[c]DR: diabetic retinopathy.

[d]NPDR: nonproliferative diabetic retinopathy.

[e]DME: diabetic macular edema.

[f]ETDRS: Early Treatment Diabetic Retinopathy Study.

[g]RWDR: referral-warranted diabetic retinopathy.

[h]STDR: sight-threatening diabetic retinopathy.

[i]PDR: proliferative diabetic retinopathy.

[j]RDR: referable diabetic retinopathy.

[k]VTDR: vision-threatening diabetic retinopathy.

Table 2. Description of smartphone ophthalmoscopy imaging techniques.

Study author, year	Attachment used	Imaging technique	Smartphone used	Ungradable
Bhat, 2016 [32]	Ocular Cellscope	Up to five fields, 50°; AI[a]**:** EyeArt v1.2 software used to grade images; acquired by: medical interns and assistants.	iPhone 5S	N/A[b]
Kim, 2017 [33]	Cellscope Retina	Both human and AI (EyeApp) graders employed.	N/A	N/A
Kim, 2018 [34]	Cellscope Retina	5-field, 50°; fields imaged: central, inferior, superior, nasal, and temporal retina; images were digitally stitched, creating a 100° image; pixels per retinal degree: 52.3; acquired by: medical students or interns.	iPhone 5S	2 (1.7%) images/eyes
Rajalakshmi, 2015 [35]	Remidio Fundus on Phone (FOP)	4-field, 45°; fields imaged: macula, disc and nasal to optic disc, superior-temporal, inferior-temporal retina; autofocus function of smartphone was used.	Android phone	0
Rajalakshmi, 2018 [36]	Remidio Fundus on Phone (FOP)	4-field, 45°; fields imaged: macula centered, disc centered, superior-temporal, and inferior-temporal retina; AI: EyeArt software used to grade images.	Android phone	5 (1.7%) patients
Russo, 2015 [37]	D-Eye (Si14 SpA, Padova, Italy)	20°; videography and digital images acquired, comprising the posterior pole, macula, optic disc, and peripheral retina; resolution: 3264×2448 pixels; pixels per retinal degree: 150; acquired by: a retinal specialist.	iPhone 5	9 (3.8%) eyes
Ryan, 2015 [38]	20 D lens	Videography and then screenshots to obtain the best images of optic nerve and macula; resolution: 3264×2488 pixels; FilmIc Pro app used to adjust focus and zoom independently; acquired by: a medical student with limited training.	iPhone 5	11 (1.8%) photographs
Sengupta, 2018 [39]	Remidio FOP	3-field, 45°; fields imaged: posterior pole (macula centered), nasal, and superotemporal field; resolution: 441 pixels per inch; acquired by: ophthalmic photographer without special training.	HTC One (M8)	1.7-2.1% of images
Toy, 2016 [40]	Volk Digital ClearField lens mounted on Paxos Scope posterior segment hardware adapter	Variable number of fields, 45°; acquired by: an ophthalmologist.	iPhone 5S	2 (2%) eyes

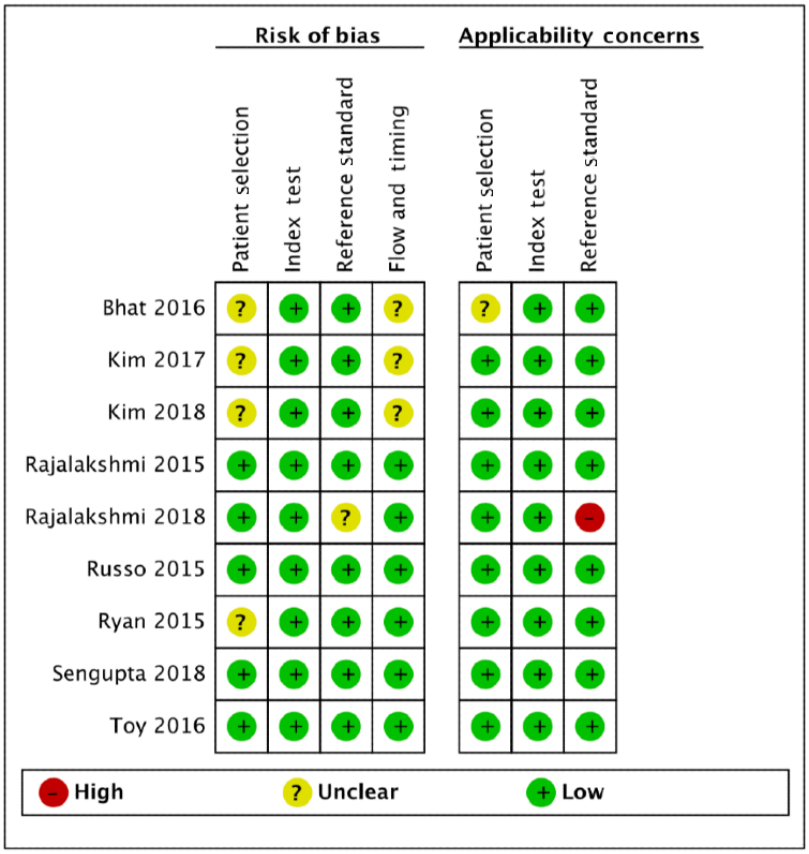
[a]AI: artificial intelligence.

[b]N/A: not available.

Quality Assessment

We carried out the quality assessment of the included studies using the QUADAS-2 criteria (Figure 2, Multimedia Appendix 4). Most studies were of high quality with low risk of bias and applicability concerns. A total of four studies had an unclear risk of bias for patient selection because of the lack of information regarding patient sampling or inappropriate exclusions. The two abstracts were of lower quality than the other studies because of the limited amount of information available. One study contained applicability concerns because it employed ophthalmologists' grading of smartphone fundoscopy images as the reference standard; it was excluded from the meta-analysis.

Figure 2. Quality of included studies assessed via Quality Assessment of Diagnostic Accuracy Studies–2 tool.



Meta-Analysis

Any Diabetic Retinopathy

In all, six studies (977 participants; Figures 3 and 4; Table 3) presented data on detecting any DR [34,35,37-40]. I^2_{LRT} was 96.8% (95% CI 94.6%-99.1%), χ^2_5=63.3, and ρ=−0.332. DOR was 100 (95% CI 27.4-368). Sensitivity and specificity ranged from 50% to 94% and 40% to 99%, respectively. The pooled sensitivity was 87.1% (95% CI 73.9%-94.2%); pooled specificity was 93.7% (95% CI 80.9%-98.1%). LR+ was 13.8 (95% CI 4.37-43.6); LR− was 0.138 (95% CI 0.066-0.287). The area under curve (AUC) was 0.957 (95% CI 0.936-0.972). Considering a pretest probability of 35.4% in diabetic patients [2], using the Fagan nomogram, the posttest probability for a positive and negative result was 88% and 7%, respectively.

We performed subgroup analysis by removing studies individually and investigating the effect on both I^2 and ρ. When one study [38] was removed, I^2 decreased to 93.0% (95% CI 86.8%-99.3%) and ρ decreased to −1.00, implying this study contributed to the heterogeneity. However, both the type of ophthalmoscopy (direct vs indirect) and reference standard used did not account for the heterogeneity.

Figure 3. Forest plot of the sensitivity and specificity of smartphone ophthalmoscopy in detecting different grades of diabetic retinopathy. AI: artificial intelligence; FN: false negatives; FP: false positives; NPDR: nonproliferative diabetic retinopathy; PDR: proliferative diabetic retinopathy; RWDR: referral-warranted diabetic retinopathy; STDR: sight-threatening diabetic retinopathy; TN: true negatives; TP: true positives; VTDR: vision-threatening diabetic retinopathy.

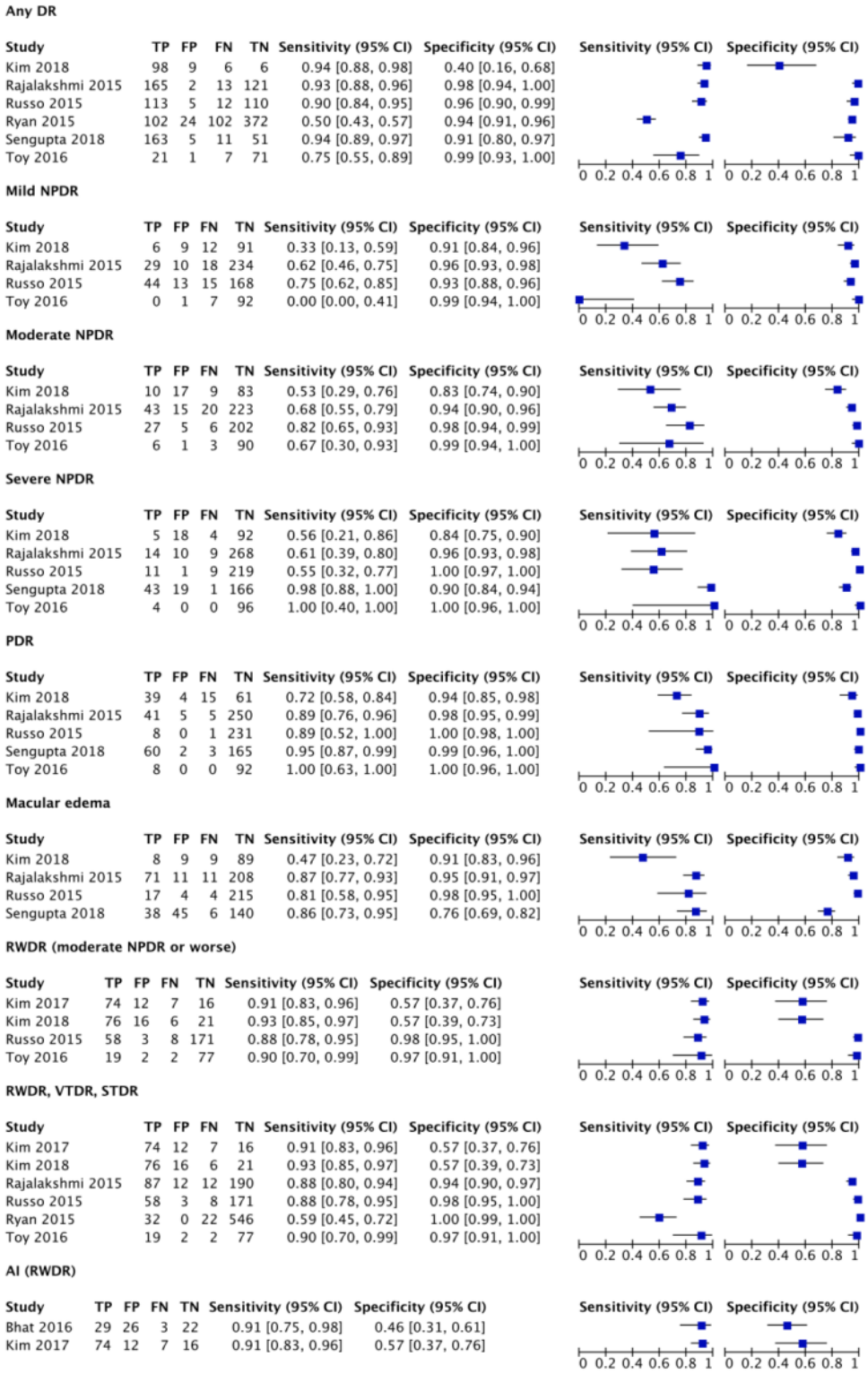


Figure 4. Summary receiver operating characteristic curves of smartphone ophthalmoscopy in detecting (A) any diabetic retinopathy; (B) mild nonproliferative diabetic retinopathy; (C) moderate nonproliferative diabetic retinopathy; (D) severe nonproliferative diabetic retinopathy; (E) proliferative diabetic retinopathy; (F) diabetic macular edema; (G) referral-warranted diabetic retinopathy, vision-threatening diabetic retinopathy, or sight-threatening diabetic retinopathy; (H) artificial intelligence to detect referral-warranted diabetic retinopathy. HSROC: hierarchical summary receiver operating characteristic.

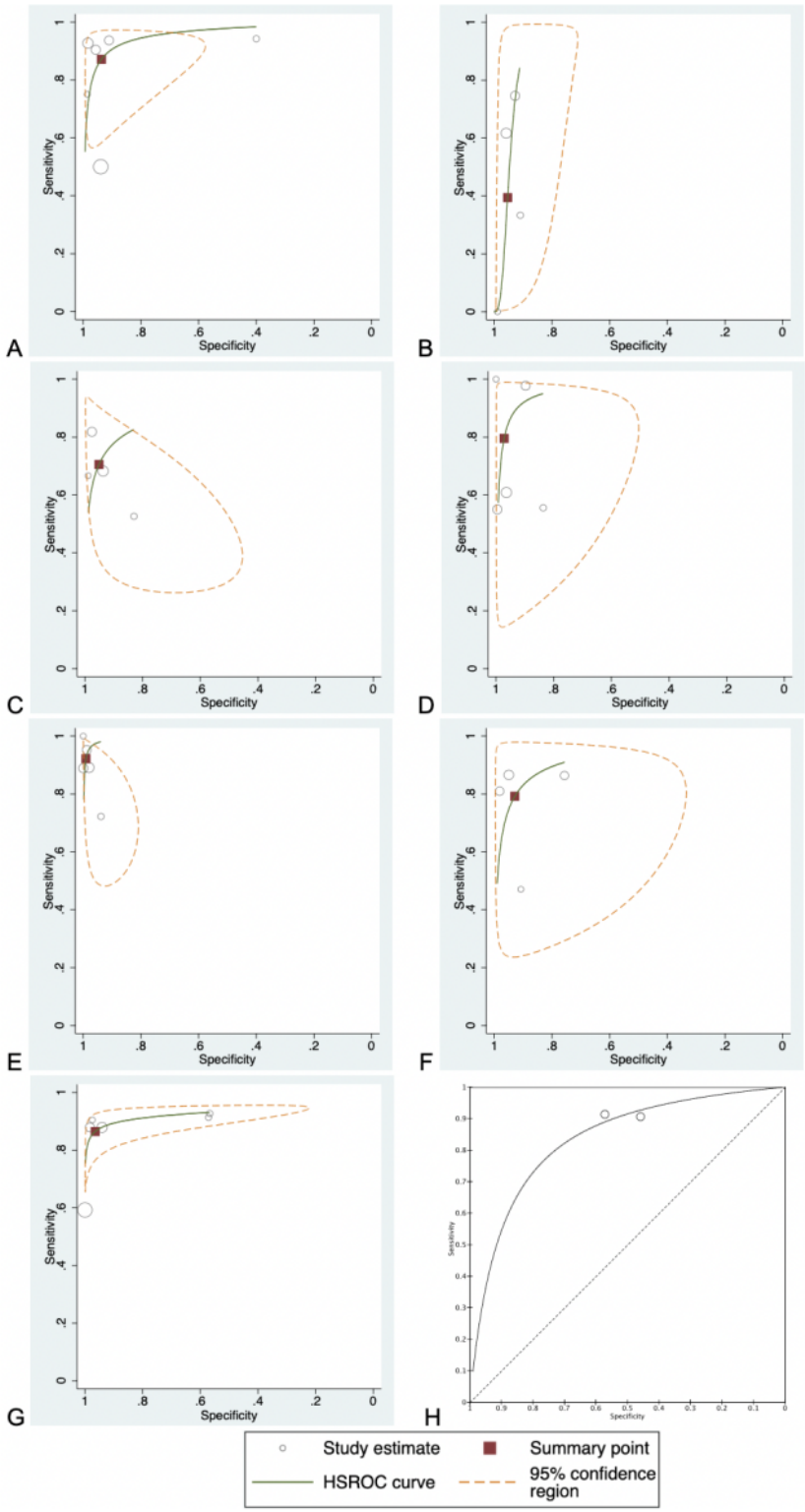


Table 3. Summary of smartphone ophthalmoscopy's test accuracy in detecting different grades of diabetic retinopathy.

DR[a] staging	Studies, n	Overall pooled sensitivity, % (95% CI)	Overall pooled specificity, % (95% CI)	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)	Diagnostic odds ratio (95% CI)	Area under summary receiver operating characteristic curve (95% CI)
Any DR	6	87 (74-94)	94 (81-98)	14 (4.4–44)	0.14 (0.06-0.29)	100 (27.4-368)	0.957 (0.936-0.972)
Mild NPDR[b]	4	39 (10-79)	95 (91-98)	8.6 (3.6-20)	0.64 (0.32-1.3)	13.6 (3.14-58.5)	0.939 (0.915-0.957)
Moderate NPDR	4	71 (57-81)	95 (88-98)	15 (4.9-43)	0.31 (0.20-0.49)	46.9 (10.6-208)	0.879 (N/A)
Severe NPDR	5	80 (49-94)	97 (88-99)	28 (6.1-133)	0.21 (0.069-0.65)	134 (17.5-1040)	0.965 (0.945-0.978)
PDR[c]	5	92 (79-97)	99 (96-99)	97 (22-425)	0.079 (0.027-0.23)	1225 (117-12,800)	0.979 (N/A)
DME[d]	4	79 (63-89)	93 (82-97)	11 (4.2-30)	0.22 (0.12-0.42)	49.8 (13.7-180)	0.925 (0.898-0.945)
RWDR[e] (moderate NPDR or worse)	4	91 (86-94)	89 (56-98)	8.1 (1.6-41)	0.11 (0.072-0.16)	75.8 (13.9-414)	0.921 (0.894-0.941)
RWDR, VTDR[f], STDR[g]	6	87 (77-92)	96 (71-99)	24 (2.6-226)	0.14 (0.087-0.23)	171 (25.9-1142)	0.929 (0.903-0.949)
AI[h] (RWDR)	2	91 (84-96)	50 (38-62)	1.8 (1.4-2.3)	0.17 (0.088-0.32)	11.3 (4.92-26.1)	N/A[i]

[a]DR: diabetic retinopathy.

[b]NPDR: nonproliferative diabetic retinopathy.

[c]PDR: proliferative diabetic retinopathy.

[d]DME: diabetic macular edema.

[e]RWDR: referral-warranted diabetic retinopathy.

[f]VTDR: vision-threatening diabetic retinopathy.

[g]STDR: sight-threatening diabetic retinopathy.

[h]AI: artificial intelligence.

[i]N/A: not available.

Mild Nonproliferative Diabetic Retinopathy

In all, four studies (542 participants) presented data on detecting mild NPDR [34,35,37,40]. I^2_{LRT} was 81.5% (95% CI 60.6%-100%), χ^2_3=10.8, and ρ=−0.862. DOR was 13.6 (95% CI 3.14-58.5). Sensitivity and specificity ranged from 0% to 75% and 91% to 99%, respectively. The pooled sensitivity was 39.4% (95% CI 10.1%-79.0%); pooled specificity was 95.4% (95% CI 91.3%-97.6%). LR+ was 8.60 (95% CI 3.64-20.3); LR− was 0.635 (95% CI 0.323-1.25). One study [40] using a lens mounted on Paxos scope yielded a sensitivity of 0%. The AUC was 0.939 (95% CI 0.915-0.957).

Moderate Nonproliferative Diabetic Retinopathy

A total of four studies (542 participants) presented data on detecting moderate NPDR [34,35,37,40]. DOR was 46.9 (95% CI 10.6-208; I^2_{DOR}=85.4%; χ^2_3=20.6). Sensitivity and specificity ranged from 53% to 82% and 83% to 99%, respectively. The pooled sensitivity was 70.5% (95% CI 56.6%-81.4%); pooled specificity was 95.1% (95% CI 87.8%-98.2%). LR+ was 14.5 (95% CI 4.89-43.2); LR− was 0.310 (95% CI 0.195-0.492). The AUC was approximately 0.879.

One study [39] assessed the sensitivity and specificity of smartphone ophthalmology in detecting R1 disease (ie, mild and moderate NPDR) to be 88.2% (95% CI 85.7%-91.6%) and 83.4% (95% CI 78%-87%), respectively.

Severe Nonproliferative Diabetic Retinopathy

Overall, five studies (677 participants) presented data on detecting severe NPDR [34,35,37,39,40]. I^2_{LRT} was 94.0% (95% CI 88.9%-99.1%), χ^2_4=33.4, and ρ=−0.111. DOR was 134 (95% CI 17.5-1039). Sensitivity and specificity ranged from 55% to 100% and 84% to 100%, respectively. The pooled sensitivity was 79.5% (95% CI 48.6%-94.1%); pooled specificity was 97.1% (95% CI 87.7%-99.4%). LR+ was 28.4 (95% CI 6.06-133); LR− was 0.211 (95% CI 0.0688-0.645). The AUC was 0.965 (95% CI 0.945-0.978).

Removing one study [34] employing medical students and interns for smartphone ophthalmoscopy led to the greatest decrease in ρ to −0.639, indicating that the remaining studies fitted well within the SROC curve. However, removing the study utilizing indirect ophthalmoscopy [40] resulted in both a decrease in ρ and I^2 to −0.464 and 93.8% (95% CI 88.5%-99.2%), respectively. Thus, our subgroup analysis for *severe NPDR* was inconclusive.

Proliferative Diabetic Retinopathy

A total of five studies (677 participants) presented data on detecting PDR [34,35,37,39,40]. DOR was 1225 (95% CI 117-12,800; I^2_{DOR}=78.0%; χ^2_4=18.2). Sensitivity and specificity ranged from 72% to 100% and 94% to 100%. The pooled sensitivity was 92.1% (95% CI 79.1%-97.4%); pooled specificity was 99.0% (95% CI 96.1%-99.8%). LR+ was 96.6 (95% CI 21.9-425); LR− was 0.0789 (95% CI 0.0273-0.228). The AUC was approximately 0.979.

Removing one study [34] decreased I^2_{DOR} to 0.0%. This study employed a medical student and an intern to acquire smartphone ophthalmoscopy images, potentially resulting in heterogeneity. Removing the only study [35] using 7-field ETDRS fundus photography as a reference standard, or another study [40] utilizing indirect ophthalmoscopy, did not reduce I^2_{DOR}.

Diabetic Macular Edema

Although the diagnosis of DME generally requires stereoscopic retinal imaging, these studies used substitute markers, such as the presence of hard exudates or laser photocoagulation scars.

In all, four studies (627 participants) presented data on detecting DME [34,35,37,39]. I^2_{LRT} was 87.9% (95% CI 75.5%-100%), χ^2_3=16.6, and ρ=−0.038. DOR was 49.8 (95% CI 13.7–180). Sensitivity and specificity ranged from 47% to 87% and 76% to 98%, respectively. The pooled sensitivity was 79.2% (95% CI 63.2%-89.4%); pooled specificity was 92.9% (95% CI 82.3%-97.4%). LR+ was 11.1 (95% CI 4.22-29.5); LR− was 0.224 (95% CI 0.119-0.422). The AUC was 0.925 (95% CI 0.898–0.945). Considering a pretest probability of 7.48% in diabetic patients [2], using the Fagan nomogram, the posttest probability for a positive and negative result was 47% and 2%, respectively.

Referral-Warranted Diabetic Retinopathy

In all, four studies (313 participants) presented data on detecting RWDR [33,34,37,40]. I^2_{LRT} was 94.3% (95% CI 89.6%-99.1%), χ^2_3=35.3, and ρ=−1.00. DOR was 75.8 (95% CI 13.9-414). Sensitivity and specificity ranged from 88% to 93% and 57% to 98%, respectively. The pooled sensitivity was 90.5% (95% CI 85.5%-93.8%); pooled specificity was 88.9% (95% CI 56.2%-98.0%). LR+ was 8.13 (95% CI 1.63-40.5); LR− was 0.107 (95% CI 0.0721-0.159). The AUC was 0.921 (95% CI 0.894-0.941).

Referral-Warranted Diabetic Retinopathy, Vision-Threatening Diabetic Retinopathy, and Sight-Threatening Diabetic Retinopathy

Overall, six studies (914 participants) presented data on detecting RWDR, VTDR, and STDR [33-35,37,38,40]. I^2_{LRT} was 98.6% (95% CI 97.8%-99.4%), χ^2_5=139, and ρ=−1.00. DOR was 171 (95% CI 25.9-1142). Sensitivity and specificity ranged from 59% to 93% and 57% to 100%, respectively. The pooled sensitivity was 86.5% (95% CI 77.1%-92.4%); pooled specificity was 96.4% (95% CI 71.1%-99.7%). LR+ was 24.1 (95% CI 2.58-226); LR− was 0.140 (95% CI 0.0865-0.228). The AUC was 0.929 (95% CI 0.903-0.949). Owing to a good fit of the SROC curve, subgroup analysis was not performed.

One study excluded from the analysis found the agreement for detecting VTDR to be high, κ=0.76 (95% CI 0.68-0.85) [39].

Artificial Intelligence in Smartphone Ophthalmoscopy

In all, two studies (152 participants) presented data on detecting RWDR using AI to grade retinal images acquired via smartphone ophthalmoscopy compared with conventional slit-lamp biomicroscopy [32,33]. I^2_{LRT} was 0.0%. DOR was 11.3 (95% CI 4.92-26.1). Specificity ranged from 46% to 57%. The pooled sensitivity was 91.2% (95% CI 84.3%-95.7%); pooled specificity was 50.0% (95% CI 38.3%-61.7%). LR+ was 1.80 (95% CI 1.42-2.28); LR− was 0.167 (95% CI 0.088-0.316). Owing to the limited number of included studies, the fixed effects model Moses-Littenberg SROC curve was employed for this analysis, and a 95% confidence region was not available.

Another study (301 participants) compared an AI's grading of smartphone ophthalmoscopy images with the reference standard ophthalmologists' grading of the same images [36]. It reported a high sensitivity of 95.8% (95% CI 92.9%-98.7%), 99.1% (95% CI 95.1%-99.9%), and 99.3% (95% CI 96.1%-99.9%), and a specificity of 80.2% (95% CI 72.6%-87.8%), 80.4% (95% CI 73.9%-85.9%), and 68.8% (95% CI 61.5%-76.2%) for any DR, STDR, and RWDR, respectively.

Discussion

Summary of Results

Overall, smartphone ophthalmoscopy performed well in detecting DR. Depending on the severity of DR, smartphone ophthalmoscopy had different accuracy. Progressing from mild NPDR to PDR, we observed an increasing trend in smartphone ophthalmoscopy's sensitivity, specificity, and DOR. In addition, smartphone ophthalmoscopy had the best performance in detecting PDR, RWDR, VTDR, and STDR; these are important categories to detect as they can significantly affect vision. The lowest sensitivity was observed for detecting mild NPDR, mainly caused by one study enrolling only 7 participants with RWDR. The DOR was lowest for AI's detection of RWDR. There was also a low percentage of ungradable images across most studies, implying that smartphone ophthalmoscopy is relatively reliable. Common causes of ungradable images included cataracts, poor pupil dilation, vitreous hemorrhages, or poor image focus.

Most studies performed smartphone direct ophthalmoscopy utilizing one of four different attachments. The included studies also assessed two methods of indirect ophthalmoscopy. Smartphone ophthalmoscopy in the included studies surpasses the UK NHS targets requiring DR retinal imaging equipment to have a minimum resolution of 6 megapixels or 30 pixels per retinal degree [34,37]. Two studies [35,36] assigned DR grades at a patient level instead of assessing each eye individually. In some cases, smartphone apps were used to digitally stitch the multiple images obtained per eye or enhance the image acquisition process by facilitating ergonomic focusing and capturing of images. Videography was used in two studies

[37,38]. Different grading criteria and reference standards were applied across the included studies. High heterogeneity among studies was observed for most types of DR—especially for *moderate NPDR* and *PDR*—except for studies employing AI to detect RWDR. In other studies reporting on the use of non-AI smartphone ophthalmoscopy in *mild NPDR, RWDR only,* and *RWDR, VTDR, and STDR*, a significant proportion of heterogeneity can be attributed to the threshold effect.

The diagnostic accuracy of AI in grading smartphone ophthalmoscopy images was unexpectedly low. In two studies, the specificity and DOR of AI in detecting RWDR was lower than that of human graders (retinal specialists and ophthalmologists). Nevertheless, one of those studies employed both human and AI to grade identical smartphone-acquired images; the specificity of AI was higher than that of humans. In contrast, a 2015 study demonstrated that AI detects RWDR in smartphone ophthalmoscopy images with 100% sensitivity and 80% specificity (AUC 0.94) [41]. In addition, a recent review revealed that AI software achieved high sensitivity and specificity for detecting DR in datasets of fundus images acquired from other imaging modalities [42]. Finally, IDx-DR was the first commercially approved AI-based autonomous diagnostic system for DR detection. In a prospective study of 900 participants, this system attained high sensitivity and specificity of 87.2% (95% CI 81.8%-91.2%) and 90.7% (95% CI 88.3%-92.7%), respectively, in detecting more than mild DR [43].

Smartphone ophthalmoscopy is a safe means of acquiring retinal images [44]. One study [34] surveyed patients on their comfort levels while undergoing retinal imaging and revealed that all participants felt more comfortable with the light from Cellscope Retina than the light from slit lamps. Other studies [38,39] employing either an intrinsic smartphone light source or external light sources reported lower luminance than conventional fundus cameras.

Comparison to Existing Studies

To our knowledge, this is the first meta-analysis evaluating the diagnostic accuracy of smartphone ophthalmoscopy for detecting DR in diabetic patients. A meta-analysis [45] evaluated the agreement between smartphone retinal imaging and retinal cameras encompassing multiple eye pathologies such as DR, glaucoma, and ocular hypertension. It reported excellent image quality in 84.7% of smartphone images, with good diagnostic accuracy; combined κ agreement was 77.8% (95% CI 70.34%-83.70%), AUC=0.86. However, the patient selection was not limited to diabetic individuals. A large study [46] involving 1460 participants (2920 eyes) had previously evaluated the diagnostic accuracy of smartphone ophthalmoscopy for optic disc imaging. Videography was performed with Peek Retina adaptor attached to an 8.0-megapixel Samsung SIII smartphone. This technique demonstrated excellent agreement (weighted κ=0.69) with a reference standard tabletop fundus camera in measuring vertical cup-disc ratio. Using smartphone ophthalmoscopy, 79.5% of eyes were gradable, compared with 86.4% for tabletop retinal imaging. Furthermore, there was no significant difference between image quality acquired by professional and inexperienced photographers. This study reported a lower percentage of gradable eyes compared with most studies included in our scoping review. This could be attributed to inherent differences in the process of DR grading (which requires examination of the retina in general) compared with measuring cup-disc ratio (which specifically examines the optic disc). Regardless of sample size, the agreement of smartphone ophthalmoscopy with a well-established reference standard remains high.

Strengths

This scoping review aimed to provide a comprehensive analysis of the available literature in this field. Correspondingly, we had broad inclusion criteria encompassing different smartphone ophthalmoscopy techniques, reference standards, DR severity scales, and health care professionals. Smartphone retinal imaging is an emerging technology, and we wanted to capture as much of the available evidence as possible (Multimedia Appendix 5). The included studies were published relatively recently, from 2015 to 2018, heralding future breakthroughs in the diagnostic accuracy of smartphone retinal imaging as affordable and accessible means of DR detection.

Our study employed a comprehensive search strategy and examined studies from different countries involving different types of diabetic patients. At least two reviewers performed quality assessment and data extraction independently following Cochrane methodology. Based on the QUADAS-2 tool, most studies possessed minimal risk of bias and little applicability concerns. In particular, all included studies blinded or masked the graders.

Limitations

Although the protocol for this scoping review was published in *BMJ Open*, this protocol was not registered. Three studies [33,34,39] utilized two graders, thereby creating two separate 2×2 tables; to avoid double-counting, we averaged the TP, FP, TN, and FN values, and rounded those average values to the nearest whole number for analysis. For one study [39], we assumed none of the four excluded eyes had DME. Owing to the small number of studies and limited information available, we were not able to conduct a meta-regression analysis or assess for publication bias.

The large 95% CIs for most SROC curves indicate imprecision. Although only studies involving diabetic patients were included, most studies were conducted in tertiary health care settings: eye or diabetes clinics. These settings can afford tabletop or portable fundus cameras. Instead, smartphone ophthalmoscopy is more relevant for screening in primary settings or resource-constrained countries. All studies required mydriasis, despite the availability of nonmydriatic smartphone ophthalmoscopy attachments [17].

Implications for Future Research

Future studies on smartphone ophthalmoscopy could utilize more consistent reference standards, such as the gold standard 7-field ETDRS stereoscopic color photographs, and standardize the DR classification criteria. Such standardization minimizes bias and heterogeneity between studies. In addition, ultrawide-field retinal imaging may detect DR features outside

the 7-field ETDRS field of view, which may be of clinical significance [47]. More studies could focus on (1) indirect ophthalmoscopy or ultrawide-field retinal imaging; (2) nonmydriatic techniques; (3) AI; and (4) primary health care settings where the comorbidities and prevalence of DR in this demographic differs.

Conclusions

Smartphone ophthalmoscopy may have an important role in identifying DR in areas with limited access to expensive retinal imaging equipment and trained staff. Our findings show that smartphone ophthalmoscopy performs well in detecting DR. However, the included studies were scarce and heterogeneous and provided imprecise findings. Future studies should use more consistent reference standards and DR classification criteria, evaluate other available forms of smartphone ophthalmoscopy, and recruit participants from primary care settings.

Acknowledgments

The authors thank Ms Soong Ai Jia for her contribution to the data extraction stage of this review. The authors gratefully acknowledge funding support from Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore.

Authors' Contributions

LTC conceived the idea for this study. CHT and BK screened the articles, extracted the data, and performed the analysis. CHT and LTC wrote the manuscript. BK, HS, CT, and LTC revised the manuscript critically.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Medical Literature Analysis and Retrieval System Online (MEDLINE), EMBASE, and Cochrane Library search strategy.
[PDF File (Adobe PDF File), 54 KB - jmir_v22i5e16658_app1.pdf]

Multimedia Appendix 2
Data extraction sheet.
[PDF File (Adobe PDF File), 70 KB - jmir_v22i5e16658_app2.pdf]

Multimedia Appendix 3
Additional details of included studies.
[PDF File (Adobe PDF File), 58 KB - jmir_v22i5e16658_app3.pdf]

Multimedia Appendix 4
Details of quality assessment of included studies using Quality Assessment of Diagnostic Accuracy Studies-2.
[PDF File (Adobe PDF File), 38 KB - jmir_v22i5e16658_app4.pdf]

Multimedia Appendix 5
Additional study details received from study investigators.
[PDF File (Adobe PDF File), 173 KB - jmir_v22i5e16658_app5.pdf]

References

1. Cheung N, Mitchell P, Wong TY. Diabetic retinopathy. Lancet 2010 Jul 10;376(9735):124-136. [doi: 10.1016/S0140-6736(09)62124-3] [Medline: 20580421]
2. Lee R, Wong TY, Sabanayagam C. Epidemiology of diabetic retinopathy, diabetic macular edema and related vision loss. Eye Vis (Lond) 2015;2:17 [FREE Full text] [doi: 10.1186/s40662-015-0026-2] [Medline: 26605370]
3. Rodrigues GB, Abe RY, Zangalli C, Sodre SL, Donini FA, Costa DC, et al. Neovascular glaucoma: a review. Int J Retina Vitreous 2016;2:26 [FREE Full text] [doi: 10.1186/s40942-016-0051-x] [Medline: 27895936]
4. Cohen S, Gardner T. Diabetic retinopathy and diabetic macular edema. Dev Ophthalmol 2016;55:137-146 [FREE Full text] [doi: 10.1159/000438970] [Medline: 26501152]
5. Stewart MW. Treatment of diabetic retinopathy: Recent advances and unresolved challenges. World J Diabetes 2016 Aug 25;7(16):333-341 [FREE Full text] [doi: 10.4239/wjd.v7.i16.333] [Medline: 27625747]
6. Mohamed Q, Gillies MC, Wong TY. Management of diabetic retinopathy: a systematic review. J Am Med Assoc 2007 Aug 22;298(8):902-916. [doi: 10.1001/jama.298.8.902] [Medline: 17712074]

7. Ellis D, Burgess PI, Kayange P. Management of diabetic retinopathy. Malawi Med J 2013 Dec;25(4):116-120 [FREE Full text] [Medline: 24926400]
8. Falavarjani KG, Wang K, Khadamy J, Sadda SR. Ultra-wide-field imaging in diabetic retinopathy; an overview. J Curr Ophthalmol 2016 Jun;28(2):57-60 [FREE Full text] [doi: 10.1016/j.joco.2016.04.001] [Medline: 27331147]
9. Panwar N, Huang P, Lee J, Keane PA, Chuan TS, Richhariya A, et al. Fundus Photography in the 21st Century--a review of recent technological advances and their implications for worldwide healthcare. Telemed J E Health 2016 Mar;22(3):198-208 [FREE Full text] [doi: 10.1089/tmj.2015.0068] [Medline: 26308281]
10. Shields CL, Materin M, Shields JA. Panoramic imaging of the ocular fundus. Arch Ophthalmol 2003 Nov;121(11):1603-1607. [doi: 10.1001/archopht.121.11.1603] [Medline: 14609919]
11. Lira RP, Oliveira CL, Marques MV, Silva AR, Pessoa CD. Adverse reactions of fluorescein angiography: a prospective study. Arq Bras Oftalmol 2007;70(4):615-618 [FREE Full text] [doi: 10.1590/s0004-27492007000400011] [Medline: 17906757]
12. Ahmed J, Ward TP, Bursell S, Aiello LM, Cavallerano JD, Vigersky RA. The sensitivity and specificity of nonmydriatic digital stereoscopic retinal imaging in detecting diabetic retinopathy. Diabetes Care 2006 Oct;29(10):2205-2209. [doi: 10.2337/dc06-0295] [Medline: 17003294]
13. Goh JK, Cheung CY, Sim SS, Tan PC, Tan GS, Wong TY. Retinal imaging techniques for diabetic retinopathy screening. J Diabetes Sci Technol 2016 Feb 1;10(2):282-294 [FREE Full text] [doi: 10.1177/1932296816629491] [Medline: 26830491]
14. Massin P, Girach A, Erginay A, Gaudric A. Optical coherence tomography: a key to the future management of patients with diabetic macular oedema. Acta Ophthalmol Scand 2006 Aug;84(4):466-474 [FREE Full text] [doi: 10.1111/j.1600-0420.2006.00694.x] [Medline: 16879566]
15. Hong SC. 3D printable retinal imaging adapter for smartphones could go global. Graefes Arch Clin Exp Ophthalmol 2015 Oct;253(10):1831-1833. [doi: 10.1007/s00417-015-3017-z] [Medline: 25900816]
16. Jones S, Edwards RT. Diabetic retinopathy screening: a systematic review of the economic evidence. Diabet Med 2010 Mar;27(3):249-256. [doi: 10.1111/j.1464-5491.2009.02870.x] [Medline: 20536486]
17. Micheletti JM, Hendrick AM, Khan FN, Ziemer DC, Pasquel FJ. Current and next generation portable screening devices for diabetic retinopathy. J Diabetes Sci Technol 2016 Feb 16;10(2):295-300 [FREE Full text] [doi: 10.1177/1932296816629158] [Medline: 26888973]
18. Fenner BJ, Wong RL, Lam W, Tan GS, Cheung GC. Advances in retinal imaging and applications in diabetic retinopathy screening: a review. Ophthalmol Ther 2018 Dec;7(2):333-346 [FREE Full text] [doi: 10.1007/s40123-018-0153-7] [Medline: 30415454]
19. Bolster NM, Giardini ME, Bastawrous A. The diabetic retinopathy screening workflow: potential for smartphone imaging. J Diabetes Sci Technol 2015 Nov 23;10(2):318-324 [FREE Full text] [doi: 10.1177/1932296815617969] [Medline: 26596630]
20. McInnes MD, Moher D, Thombs BD, McGrath TA, Bossuyt PM, the PRISMA-DTA Group, et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: The PRISMA-DTA Statement. J Am Med Assoc 2018 Jan 23;319(4):388-396. [doi: 10.1001/jama.2017.19163] [Medline: 29362800]
21. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. Hoboken, New Jersey: Wiley; 2011.
22. 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]
23. 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]
24. Tan CH, Quah W, Tan CS, Smith H, Tudor Car L. Use of smartphones for detecting diabetic retinopathy: a protocol for a scoping review of diagnostic test accuracy studies. BMJ Open 2019 Dec 8;9(12):e028811 [FREE Full text] [doi: 10.1136/bmjopen-2018-028811] [Medline: 31818832]
25. Covidence - Better systematic review management. URL: https://www.covidence.org/home [accessed 2020-03-16]
26. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011 Oct 18;155(8):529-536. [doi: 10.7326/0003-4819-155-8-201110180-00009] [Medline: 22007046]
27. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol 2003 Nov;56(11):1129-1135. [doi: 10.1016/s0895-4356(03)00177-x] [Medline: 14615004]
28. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 2005 Oct;58(10):982-990. [doi: 10.1016/j.jclinepi.2005.02.022] [Medline: 16168343]
29. Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA. A unification of models for meta-analysis of diagnostic accuracy studies. Biostatistics 2007 Apr;8(2):239-251. [doi: 10.1093/biostatistics/kxl004] [Medline: 16698768]
30. Cochrane Community. Copenhagen: The Nordic Cochrane Centre: The Cochrane Collaboration; 2014. Review Manager (RevMan) [Computer program]. Version 5.3 URL: https://community.cochrane.org/help/tools-and-software/revman-5 [accessed 2020-03-16]

31. Zamora J, Abraira V, Muriel A, Khan K, Coomarasamy A. Meta-DiSc: a software for meta-analysis of test accuracy data. BMC Med Res Methodol 2006 Jul 12;6:31 [FREE Full text] [doi: 10.1186/1471-2288-6-31] [Medline: 16836745]
32. Bhat S, Bhaskaranand M, Ramachandra C, Qi O, Liu L, Apte R, et al. Automated image analysis for Diabetic Retinopathy Screening with iPhone-based fundus camera. Invest Ophthalmol Vis Sci 2016 Sep;57(12):5964 [FREE Full text]
33. Kim T, Li P, Niziol LM, Bhaskaranand M, Bhat S, Ramachandra C, et al. Comparison of automated and expert human grading of diabetic retinopathy using smartphone-based retinal photography. Invest Ophthalmol Vis Sci 2017;58(8):659 [FREE Full text]
34. Kim TN, Myers F, Reber C, Loury PJ, Loumou P, Webster D, et al. A smartphone-based tool for rapid, portable, and automated wide-field retinal imaging. Transl Vis Sci Technol 2018 Sep;7(5):21 [FREE Full text] [doi: 10.1167/tvst.7.5.21] [Medline: 30280006]
35. Rajalakshmi R, Arulmalar S, Usha M, Prathiba V, Kareemuddin KS, Anjana RM, et al. Validation of smartphone based retinal photography for diabetic retinopathy screening. PLoS One 2015;10(9):e0138285 [FREE Full text] [doi: 10.1371/journal.pone.0138285] [Medline: 26401839]
36. Rajalakshmi R, Subashini R, Anjana RM, Mohan V. Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence. Eye (Lond) 2018 Jun;32(6):1138-1144 [FREE Full text] [doi: 10.1038/s41433-018-0064-9] [Medline: 29520050]
37. Russo A, Morescalchi F, Costagliola C, Delcassi L, Semeraro F. Comparison of smartphone ophthalmoscopy with slit-lamp biomicroscopy for grading diabetic retinopathy. Am J Ophthalmol 2015 Feb;159(2):360-4.e1. [doi: 10.1016/j.ajo.2014.11.008] [Medline: 25447109]
38. Ryan ME, Rajalakshmi R, Prathiba V, Anjana RM, Ranjani H, Narayan KM, et al. Comparison Among Methods of Retinopathy Assessment (CAMRA) Study: smartphone, nonmydriatic, and mydriatic photography. Ophthalmology 2015 Oct;122(10):2038-2043 [FREE Full text] [doi: 10.1016/j.ophtha.2015.06.011] [Medline: 26189190]
39. Sengupta S, Sindal MD, Baskaran P, Pan U, Venkatesh R. Sensitivity and specificity of smartphone-based retinal imaging for diabetic retinopathy: a comparative study. Ophthalmol Retina 2019 Feb;3(2):146-153. [doi: 10.1016/j.oret.2018.09.016] [Medline: 31014763]
40. Toy BC, Myung DJ, He L, Pan CK, Chang RT, Polkinhorne A, et al. Smartphone-based dilated fundus photography and near visual acuity testing as inexpensive screening tools to detect referral warranted diabetic eye disease. Retina 2016 May;36(5):1000-1008. [doi: 10.1097/IAE.0000000000000955] [Medline: 26807627]
41. Bhat S, Bhaskaranand M, Ramachandra C, Margolis T, Fletcher D, Solanki K. Fully-automated diabetic retinopathy screening using cellphone-based cameras. Invest Ophthalmol Vis Sci 2015;56(7):1428 [FREE Full text]
42. Ting DS, Pasquale LR, Peng L, Campbell JP, Lee AY, Raman R, et al. Artificial intelligence and deep learning in ophthalmology. Br J Ophthalmol 2019 Feb;103(2):167-175 [FREE Full text] [doi: 10.1136/bjophthalmol-2018-313173] [Medline: 30361278]
43. Abràmoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. NPJ Digit Med 2018;1:39 [FREE Full text] [doi: 10.1038/s41746-018-0040-6] [Medline: 31304320]
44. Kim DY, Delori F, Mukai S. Smartphone photography safety. Ophthalmology 2012 Oct;119(10):2200-1; author reply 2201. [doi: 10.1016/j.ophtha.2012.05.005] [Medline: 23034305]
45. Vilela MA, Valença FM, Barreto PK, Amaral CE, Pellanda LC. Agreement between retinal images obtained via smartphones and images obtained with retinal cameras or fundoscopic exams - systematic review and meta-analysis. Clin Ophthalmol 2018;12:2581-2589. [doi: 10.2147/opth.s182022]
46. Bastawrous A, Giardini ME, Bolster NM, Peto T, Shah N, Livingstone IA, et al. Clinical validation of a smartphone-based adapter for optic disc imaging in Kenya. JAMA Ophthalmol 2016 Feb;134(2):151-158 [FREE Full text] [doi: 10.1001/jamaophthalmol.2015.4625] [Medline: 26606110]
47. Silva PS, Cavallerano JD, Haddad NM, Kwak H, Dyer KH, Omar AF, et al. Peripheral lesions identified on ultrawide field imaging predict increased risk of diabetic retinopathy progression over 4 years. Ophthalmology 2015 May;122(5):949-956. [doi: 10.1016/j.ophtha.2015.01.008] [Medline: 25704318]

Abbreviations

AI: artificial intelligence
AUC: area under curve
DM: diabetes mellitus
DME: diabetic macular edema
DOR: diagnostic odds ratio
DR: diabetic retinopathy
ETDRS: Early Treatment Diabetic Retinopathy Study
FN: false negatives
FP: false positives

LR−: negative likelihood ratio
LR+: positive likelihood ratio
LRT: likelihood ratio test
NHS: National Health Service
NPDR: nonproliferative diabetic retinopathy
PDR: proliferative diabetic retinopathy
QUADAS: Quality Assessment of Diagnostic Accuracy Studies
RWDR: referral-warranted diabetic retinopathy
SROC: summary receiver operating characteristic
STDR: sight-threatening diabetic retinopathy
TN: true negatives
TP: true positives
VTDR: vision-threatening diabetic retinopathy

Edited by G Eysenbach; submitted 11.10.19; peer-reviewed by M Tse, R Raman, G Lim; comments to author 16.01.20; revised version received 12.02.20; accepted 16.02.20; published 15.05.20.

Please cite as:
Tan CH, Kyaw BM, Smith H, Tan CS, Tudor Car L
Use of Smartphones to Detect Diabetic Retinopathy: Scoping Review and Meta-Analysis of Diagnostic Test Accuracy Studies
J Med Internet Res 2020;22(5):e16658
URL: http://www.jmir.org/2020/5/e16658/
doi:10.2196/16658
PMID:32347810

Original Paper

Enhancing Patient Activation and Self-Management Activities in Patients With Type 2 Diabetes Using the US Department of Defense Mobile Health Care Environment: Feasibility Study

Ronald W Gimbel[1], PhD; Lior M Rennert[1], PhD; Paul Crawford[2], MD; Jeanette R Little[3], MS; Khoa Truong[1], MA, PhD; Joel E Williams[1], MPH, PhD; Sarah F Griffin[1], MPH, PhD; Lu Shi[1], PhD; Liwei Chen[4], MD, PhD; LingLing Zhang[5], ScD; Jennie B Moss[2], BSN, MS; Robert C Marshall[6], MPH, MD; Karen W Edwards[1], MS; Kristy J Crawford[2], BSN; Marie Hing[7], MS; Amanda Schmeltz[3], MEd; Brandon Lumsden[1], MS; Morgan Ashby[1], MS; Elizabeth Haas[1], BS; Kelly Palazzo[1], BS

[1]Department of Public Health Sciences, Clemson University, Clemson, SC, United States

[2]Nellis Family Medicine Residency Program, Mike O'Callaghan Federal Hospital, Las Vegas, NV, United States

[3]Mobile Health Innovation Center, Telemedicine & Advanced Technologies Research Center, U.S. Army Medical Research & Materials Command, Fort Gordon, GA, United States

[4]Fielding School of Public Health, University of California Los Angeles, Los Angeles, CA, United States

[5]College of Nursing and Health Sciences, University of Massachusetts Boston, Boston, MA, United States

[6]Clinical Informatics Fellowship Program, Madigan Army Medical Center, Tacoma, WA, United States

[7]Department of Internal Medicine, Madigan Army Medical Center, Tacoma, WA, United States

Corresponding Author:
Ronald W Gimbel, PhD
Department of Public Health Sciences
Clemson University
501 Edwards Hall
Clemson, SC, 29634
United States
Phone: 1 5405223759
Email: rgimbel@clemson.edu

Abstract

Background: Past mobile health (mHealth) efforts to empower type 2 diabetes (T2D) self-management include portals, text messaging, collection of biometric data, electronic coaching, email, and collection of lifestyle information.

Objective: The primary objective was to enhance patient activation and self-management of T2D using the US Department of Defense's Mobile Health Care Environment (MHCE) in a patient-centered medical home setting.

Methods: A multisite study, including a user-centered design and a controlled trial, was conducted within the US Military Health System. Phase I assessed preferences regarding the enhancement of the enabling technology. Phase II was a single-blinded 12-month feasibility study that randomly assigned 240 patients to either the intervention (n=123, received mHealth technology and behavioral messages tailored to Patient Activation Measure [PAM] level at baseline) or the control group (n=117, received equipment but not messaging. The primary outcome measure was PAM scores. Secondary outcome measures included Summary of Diabetes Self-Care Activities (SDSCA) scores and cardiometabolic outcomes. We used generalized estimating equations to estimate changes in outcomes.

Results: The final sample consisted of 229 patients. Participants were 61.6% (141/229) male, had a mean age of 62.9 years, mean glycated hemoglobin (HbA_{1c}) of 7.5%, mean BMI of 32.7, and a mean duration of T2D diagnosis of 9.8 years. At month 12, the control group showed significantly greater improvements compared with the intervention group in PAM scores (control mean 7.49, intervention mean 1.77; *P*=.007), HbA_{1c} (control mean −0.53, intervention mean −0.11; *P*=.006), and low-density lipoprotein cholesterol (control mean −7.14, intervention mean 4.38; *P*=.01). Both groups showed significant improvement in SDSCA, BMI, waist size, and diastolic blood pressure; between-group differences were not statistically significant. Except for patients with the highest level of activation (PAM level 4), intervention group patients exhibited significant improvements in PAM scores. For patients with the lowest level of activation (PAM level 1), the intervention group showed significantly greater

improvement compared with the control group in HbA_{1c} (control mean −0.09, intervention mean −0.52; *P*=.04), BMI (control mean 0.58, intervention mean −1.22; *P*=.01), and high-density lipoprotein cholesterol levels (control mean −4.86, intervention mean 3.56; *P*<.001). Significant improvements were seen in AM scores, SDSCA, and waist size for both groups and in diastolic and systolic blood pressure for the control group; the between-group differences were not statistically significant. The percentage of participants who were engaged with MHCE for ≥50% of days period was 60.7% (68/112; months 0-3), 57.4% (62/108; months 3-6), 49.5% (51/103; months 6-9), and 43% (42/98; months 9-12).

Conclusions: Our study produced mixed results with improvement in PAM scores and outcomes in both the intervention and control groups. Structural design issues may have hampered the influence of tailored behavioral messaging within the intervention group.

Trial Registration: ClinicalTrials.gov NCT02949037; https://clinicaltrials.gov/ct2/show/NCT02949037

International Registered Report Identifier (IRRID): RR2-10.2196/resprot.6993

(J Med Internet Res 2020;22(5):e17968) doi:10.2196/17968

KEYWORDS

mHealth; diabetes mellitus; patient activation; patient-centered care; eHealth

Introduction

Background

Type 2 diabetes (T2D) is a chronic disease with high rates of disability, impaired quality of life, and premature death [1-6]. The prevalence of T2D is increasing at an alarming rate in the United States; in 2017, the estimated number of patients was 30.3 million or about 9.4% of the adult population [2,3,7]. T2D is the leading cause of blindness, nontraumatic amputations, and adult renal failure [8], and it reduces life expectancy by 5 to 10 years [2]. On average, the medical expenditure of patients with T2D is about 2.3 times greater than that of a T2D-free individual [9]. The potential for mobile health (mHealth) technologies in the care of patients with T2D and other chronic conditions to date have assessed a variety of tools and techniques. Some of the most promising tools and techniques include regular collection of biometric devices (eg, glucometers, activity monitors [10,11], SMS messaging [10,12-16], secure email communication with clinical teams, and regular reporting of quality-of-life variables [17-20]). Each of these tools, used alone or in combination, has demonstrated varying degrees of effectiveness. In this study, we sought to incorporate several of the most promising mHealth capabilities in a patient-centered medical home (PCMH) workflow.

The concept of patient activation, detailed in our protocol [21], has been demonstrated to correlate with improved clinical outcomes, increased preventative care, and overall lower health care–related cost [22-25]. Research on increased activation and improved clinical outcomes using patient portal and personal health record (PHR)-based interventions have provided mixed results [16,26-33]. It is noteworthy that early manuscripts did not provide substantial detail related to the design and navigation of portal or PHRs or whether the embedded intervention included behavioral reinforcement. Several studies have demonstrated a relationship between increased patient activation and improved clinical outcomes (eg, hypertension, smoking, BMI, and glycated hemoglobin [HbA_{1c}]) [5,16,25,34-38], whereas other studies did not achieve significant improvement in clinical outcomes [5,37,39]. There is evidence to suggest that activated patients are more likely to practice healthy behaviors related to their diet [37,40] and physical activity [37,41], better manage their T2D [37], and obtain preventive screenings [42].

Objective

The primary goal of this research was to enhance patient activation and improve self-management of T2D using the US Department of Defense's (DoD) Mobile Health Care Environment (MHCE) in a PCMH setting. We hypothesized that the MHCE intervention will lead to improvement in patient activation, increase in T2D self-care, and improvement in clinical outcomes.

Methods

Trial Design

The trial design and methods are described in detail elsewhere [21]. This was a feasibility study within the DoD's Military Health System, which included a user-centered design phase and a feasibility trial conducted at two sites. In phase I, we assessed both patient and clinician preferences regarding MHCE technology capabilities and enhancements for T2D care. The phase II research was a single-blinded (patients only) 12-month feasibility study that incorporated randomization principles. We employed a 1:1 allocation ratio between the intervention and control groups.

The study was approved by the Institutional Review Boards (IRB) of Clemson University (#IRB2015-234) and the Madigan Army Medical Center (IRB #216073). The study was registered with ClinicalTrials.gov (NCT02949037) on October 31, 2016.

Participant Enrollment

Patients were recruited from the PCMH clinic schedule, provider referrals, distributed posters and fliers, and population health databases. Potential participants were scheduled for a screening visit with a study staff member to establish eligibility, discuss and execute an informed consent document, and were administered the Patient Activation Measure (PAM) instrument. Patients' PAM levels placed them in a stratified group, where patients were randomly allocated to intervention or control groups.

Clinicians practicing in the respective PCMH sites were invited to participate by the site's principal investigator; this was a convenience sample. The clinician participants met with their senior research associate to review the IRB-approved minimal-risk information sheet. For phase II, clinicians signed an informed consent document. The clinician participants were not blinded to the study.

Inclusion and Exclusion Criteria

Inclusion criteria for patient participation in phase I or II research included men and women (1) aged 18 years or older, (2) with the ability to understand and read English, (3) enrolled for primary care at one of the target PCMH sites, and (4) with a diagnosis of T2D. In addition, in phase II, we sought to recruit a maximum of 120 participants (per PCMH), with a proportional distribution of patients with PAM levels 1 through 4; a primer on PAM levels is included elsewhere [21,43]. As this is a feasibility study, we did not derive the 120 per site recruitment numbers from power calculations.

Inclusion criteria for clinician participation in phase I or II research included (1) physicians, physician assistants, nurse practitioners, or nurses at the target PCMHs and (2) providing care for patients with T2D.

Exclusion criteria for patient participation in phase I or II research included the following: (1) pregnant women; (2) non-English-speaking patients; (3) patients receiving hospice care; (4) patients having active cancer and receiving treatment with chemotherapy or radiation therapy; (5) patients taking warfarin; (6) patients that have been a recipient of gastric bypass or a similar procedure; (7) patients having a diagnosis of uncontrolled hypothyroidism; (8) patients having known Cushing syndrome; (9) patients being treated with oral steroids; (10) patients with known liver disease; (11) patients with a current diagnosis of cognitive impairments that would interfere with the use of technology; (12) patients having congestive heart failure, in New York Heart Association functional classification III or IV; and (13) patients unable to use a mobile device due to cognitive or physical impairments during initial screening.

Exclusion criteria for clinician participation in phase I or II research included the following: (1) not affiliated with the target site and (2) not providing care for patients with T2D.

Setting and Site Selection

We sought to purposefully assess MHCE use in T2D care in two distinctly different PCMH environments and locations. The patient base included those on active duty, retirees, and dependents who have typically spent years in the military.

The Madigan Army Medical Center was the US Army's second largest military treatment facility located in Tacoma, Washington, and was a tertiary facility with a level II trauma center and robust graduate medical education programs. They served a patient base of approximately 118,000 patients; about 7500 (or >6%) were living with T2D. Of the T2D population, about 15.00% (1125/7500) were active duty members or their dependents, and about 85.00% (6375/7500) were retirees and their dependents. Over half of the patients with T2D were aged 57-76 years. The study location within the medical center was an Internal Medicine PCMH with approximately 14,300 enrolled patients supported by a staff of 77 including 12 staff physicians and 8 residents.

The Mike O'Callaghan Federal Medical Center was a federal facility in the greater Las Vegas, Nevada area, which served approximately 47,000 patients; about 4500 (>9%) were living with T2D. Of the T2D population, about 4.00% (180/4500) were active duty members or their dependents, and about 96.00% (4320/4500) were retirees and their dependents. Over 72.00% (3240/4500) of the patients with T2D were aged ≥60 years. The study location within the medical center was a Family Medicine PCMH with approximately 7500 enrolled patients supported by a staff of 62 including 9 staff physicians and 26 residents.

Description of the Mobile Health Care Environment

The DoD's MHCE system is a secure health information system designed to support health services delivery and mHealth. The MHCE meets all physical and information security mandates, as prescribed by federal law and DoD regulation, for the protection of personal health information and personally identifiable information.

Intervention Overview

The intervention has been described elsewhere [21]. This intervention enhanced MHCE in several ways. First, the intervention enhanced the capacity to include collection and visualization of data from Bluetooth-enabled medical devices. This included mapping data from device output into the MHCE, developing data visualization appropriate for mHealth and clinical care (eg, graphing outcomes and temporal trend patterns), migrating data to an analysis cell, and developing decision-support algorithms that drive safety alerts and behavioral message reinforcement. The devices used in this study included a Bluetooth-enabled scale, glucometer, blood pressure reader, and activity monitor. Second, the intervention expanded the capacity of the MHCE analysis cell to manage large amounts of data and to conduct both routine reports and research applications. Third, the intervention added patient activation and associated measurement instruments to capture baseline and ongoing changes to patient activation. Fourth, the intervention expanded the MHCE messaging platform that research associates used to send tailored behavioral messaging to patients in an effort to influence greater activation and reinforce positive behavior. An annotated visual presentation of patient screenshots and workflow is included in Multimedia Appendix 1.

The MHCE was accessed by mobile phones and tablets that used either an IOS or Android platform. The MHCE requires internet access for patients to sync data from devices to the MHCE backend portal, to receive tailored behavioral messages, or to support other functions accessible by clinicians (Multimedia Appendix 2). During the study, patients received SMS messages with hyperlinks to a separate secure information system platform used for the administration and analysis of PAM and Summary of Diabetes Self-Care Activities (SDSCA) instruments.

Tailored Behavioral Messaging

A primary component of the MHCE system was tailored behavioral messaging. In this study, the research team developed behavioral messages tailored for each of the four PAM score levels; in total, we developed over 360 messages. The messages fell within 9 functional areas common to T2D care and the SDSCA survey: nutrition, home monitoring, physical activity, blood pressure, foot care, medications, smoking, glucose control, and general behavioral reinforcement. The message development process is outlined below (Textbox 1).

Textbox 1. Tailored behavioral messaging development process.

A multistage message development process was guided by patient activation levels and theoretical constructs relevant to those levels. Specifically:

- Step 1: Health behavior researchers developed messages tailored for each of the four Patient Activation Measure (PAM) score levels and five type 2 diabetes self-care behaviors measured by the Summary of Diabetes Self-Care Activities.
- Step 2: Messages were assessed by two-person teams for content accuracy, reading level, and message appropriateness for the PAM level. Messages were then revised based on this review process. A clinical advisory team consisting of clinicians and researchers reviewed and approved the messages for clinical relevance and content.
- Step 3: A pretest (n=21) of the messages with a population similar to those targeted for the larger study was administered. Each person was provided with a random sample of five messages based on their PAM level. They were then asked four close-ended questions and one open-ended question about the messages.
- Step 4: Researchers reviewed PAM level–specific messages and rated them as acceptable, questionable, or unacceptable based on appropriateness for PAM level, behavioral theory construct, reading level, and content accuracy. Messages were then revised based on this review process.

This process supported triangulation review of findings that yielded a set of 360 PAM-level appropriate messages. At each step in the process, messages were reviewed and revised based on research team member feedback (stages 2 and 4) or by potential study participant feedback (stage 3). Overall, the first two steps in the process produced messages with a high degree of acceptability by people very similar to the study population. Pretest participants found the messages to be encouraging, useful, applicable, and impactful. This indicates a high level of credibility with regard to the application of the theoretical constructs within the messages.

For patients with baseline PAM level 1, we used messages addressing the emotional state of feeling overwhelmed and passive with an emphasis on the importance of taking action. PAM level 2 messages were designed to build knowledge and self-efficacy and focus on taking small steps that can be accomplished without much in-depth knowledge. PAM level 3 messages focused on building self-management skills, such as goal setting and self-monitoring. PAM level 4 messages focused on avoiding relapse when stressed.

In phase II of our study, tailored behavioral messages were delivered to each intervention group participant, via the MHCE accessed through their mobile device, based on both senior research associate–initiated and algorithm-automated schedules and thresholds developed according to PAM level, SDSCA responses, and agreed-upon general rotation. Senior research associates used the MHCE backend portal control panel for manual rotational scheduling of messages delivered 3 days per week. Participant responses to the SDSCA and data from synced devices triggered additional messaging if their clinical readings from biomedical devices exceeded the established thresholds. If a patient's PAM level changed within the course of the study, the tailored behavioral messaging to that patient changed to align with the new PAM level.

Phase I User-Centered Design

In phase I, we captured feedback from patients on MHCE navigation, use of external devices, ease of use, and satisfaction. We collected baseline research participant data following informed consent. One researcher led individual participants through usability testing, and an additional researcher-observer recorded observations using an observation guide for recording time on task, number of attempts by task, task sequences, and participant questions/comments. Following a researcher-provided demonstration of the MHCE, each participant was requested to concurrently navigate each component of the MHCE system via a mini tablet device under their control. For each task, each participant was asked three open-ended questions to evaluate task-specific user satisfaction. The researcher delivered a brief demonstration of the external devices used in the study. Participants were requested to (1) manually upload data, (2) sync each device with the app, and (3) interpret graphs. The data collection instrument used in phase I is shown in Multimedia Appendix 3.

Researchers have evaluated usability by applying definitions and evaluation metrics guided by the International Organization for Standardization's 9241-11 usability framework and mHealth usability research [44]. Specific metrics to evaluate usability are effectiveness, efficiency, and satisfaction. Researchers used the Single Ease Question to evaluate informant satisfaction immediately after performing each task [45]. The system usability scale (SUS) was measured for overall informant satisfaction with the MHCE [46].

We assessed clinician preferences in phase I using focus groups at each study site. Two trained researchers facilitated focus groups using a semistructured interview guide to elicit feedback about the MHCE (Multimedia Appendix 4). A four-member team completed a thematically organized data analysis of the clinician and nurse feedback using an inductive narrative approach [47-49].

Phase II Controlled Study: Patient Enrollment and Study Flow

In phase II, we aimed to recruit 240 patients (120 per site), with one half assigned to a control group. Eligible patients were assigned to four strata according to their PAM levels. After all

patients were assigned, simple randomization was performed within each stratum to assign patients to either the MHCE (intervention) or control groups. Phase II research was conducted between July 2017 and November 2018.

Mobile Health Care Environment Intervention Versus Intervention-Lite (Control)

Patients in both the intervention and intervention-lite (control) groups received the device package mentioned above. These devices collected and recorded biometric data; participants were trained on the biomedical device and peripheral equipment use.

For the intervention group, the devices were mapped to the MHCE system accessible from the patients' mobile phone or an iPad mini tablet device. Data from their biomedical devices were visually presented in the MHCE with trend and scalable options. Safety algorithms were mapped to these clinical data to alert the participant and, depending on the measure, the clinical team when readings exceeded the established thresholds. The devices for the control group participants were not connected to the MHCE system.

Initial Outcome Measures for Patient Component

Primary outcome measures included the PAM scores. Secondary outcome measures in the study were (1) SDSCA responses, (2) clinical measures (Textbox 1), (3) comorbid conditions, and (4) SUS survey scores.

Patient Activation Measure Instrument

The self-reported PAM survey is associated with self-management behaviors, medication adherence, patient satisfaction, and quality of life [50,51]. Within a T2D-specific population, PAM is not related to knowledge regarding HbA_{1c} (the standard measure of average blood glucose level [52]), but it is associated with better glycemic control [53]. The PAM is a valid, reliable, unidimensional, and probabilistic Guttman-like scale instrument [43], and it is a standard tool to measure patient activation. We administered the PAM at screening visits in phases I and II and electronically every 3 months during phase II for both the intervention and control groups. In computing PAM in the analysis, we leveraged the 0-100 scoring mechanism based on licensing guidance from *Insignia Health*. PAM levels 1-4 are determined by the continuous PAM score.

Summary of the Diabetes Self-Care Activities Instrument

The SDSCA instrument is a brief self-report instrument for measuring levels of self-management across different components of the T2D regimen [54]. The SDSCA includes 11 core items associated with T2D self-care. The SDSCA has been successfully used in numerous T2D studies both within and outside the United States [54-58]. The SDSCA has been validated and is considered a standard instrument in T2D care for measuring self-care activities [54]. We administered the SDSCA at the intake visit for phase II and electronically every 2 weeks during phase II for both the intervention and control groups.

Clinical Measures

We collected clinical measures (Textbox 2) from patients at intake during phase I research. We collected and compared changes in clinical measures for both groups in phase II at 3 points: intake, midpoint (month six), and conclusion (month 12). Data were manually abstracted by senior research associates from the PHR. For patients assigned to the MHCE intervention group, the MHCE system recorded weight, systolic blood pressure, diastolic blood pressure, and blood glucose values to the MHCE module on a regular basis via Wi-Fi or Bluetooth-enabled peripheral equipment.

Textbox 2. Clinical measures in phase II.

- Glycated hemoglobin
- Low-density lipoprotein
- High-density lipoprotein
- BMI
- Abdominal circumference
- Systolic blood pressure
- Diastolic blood pressure

System Usability Scale Survey

The SUS survey is a 10-item Likert-like scaled survey used to convey a subjective assessment of system usability. The instrument was developed over 15 years ago and is used to measure the usability of websites [59]. In this study, we substituted the term "MHCE system" for the term "website" in the instrument. Phase I participants completed the SUS survey at the conclusion of the encounter. Phase II intervention group participants completed the SUS survey at the midpoint and study conclusion.

Comorbid Conditions

We assessed and documented comorbid conditions among both the control and intervention groups during the prescreening of eligibility, at intake, at study midpoint, and at study conclusion. Although not primary outcome measures, any change in comorbid conditions was assessed.

Data Analysis Strategy

We conducted the primary analyses for phase II using an intent-to-treat approach. Achievement of randomization was evaluated by comparing the baseline key variables between the MHCE intervention group and the control group. These baseline

characteristics were compared using Kruskal-Wallis tests for continuous variables, Cochran-Armitage trend tests for ordinal variables, and chi-squared tests for binary variables and multinomial variables. Exact tests were performed to account for categorical variables with small cell counts.

Generalized estimating equations (GEE) with an autoregressive correlation structure were used to test the hypotheses that patients who participated in MHCE had higher PAM, SDSCA, and improved selected clinical outcomes and comorbid conditions than their control counterparts. The predictors in the model included a binary variable for a group, a three- or five-level categorical variable for time (depending on the number of repeated measurements for the outcome), and their interaction. All models included covariate adjustment for site, age, sex, race, duration of disease, work status, and baseline physical activity level. As the intervention group messages were tailored to each subject based on their baseline PAM level, we examined the effect of intervention within each baseline PAM level by including a three-way interaction between intervention, time, and baseline PAM level in the GEE models. CIs are reported to assess changes within groups. Within-group changes (from baseline) in outcomes are deemed significant at the alpha=.05 level if the estimated CIs exclude 0. Between-group comparisons of change (from baseline) in outcomes are deemed significant for two-sided *P* values< 0.5.

GEE were also used to test the hypothesis that patients who engaged at a higher rate with the interactive and tailored mobile technology in MHCE realized greater improvement in PAM level, SDSCA, and clinical measures. We defined the main independent variable for MHCE usage by an indicator variable set to 1 if the percentage of days engaged with MHCE in a 3-month period exceeded 80% (strong adherence). As adherence varies with time, the GEE model assumed working independence for the correlation structure [60]. All GEE analyses were conducted using the Generalized Linear Model procedure in the Statistical Analysis System.

Results

Phase I—User-Centered Design Research Participants

The research team recruited 10 research participants per site, with 20 participants for phase I. The study sample was 70% (14/20) male, 65% (13/20) white, middle-aged (average 62.8, range 40-82 years), obese (median BMI 33.7, range 24.3-43.5), with elevated blood sugar levels (median HbA_{1c} 7.2, range 5.8-12.2), and increased patient activation, 75% (15/20) with PAM level 3 scores at intake. Phase I research was conducted between November and December 2016.

Phase I—Participant (Patients) Effectiveness With Mobile Health Care Environment Technology

Overall, participants performed better with goal-setting tasks (almost all completed these tasks without assistance), showed some difficulty with tasks that required manual entry (the majority could do so with little to no assistance), and had the most difficulty with Bluetooth syncing of devices (approximately half needed assistance or could not complete; Table 1).

Table 1. Phase I participant effectiveness with the technology. Here, n denotes the number of responses in each category.

Site and effectiveness	Blood pressure goal setting	Glucose goal setting	Weight goal setting	Steps goal setting	BP manual entry	BP sync with device	Glucose manual entry	Weight manual entry	Weight sync with device
Madigan, n									
Completed without help	9	9	9	5	7	1	7	5	4
Needed hints	1	0	1	4	3	7	3	5	3
Failed to complete	0	1	0	1	0	2	0	0	3
Nellis, n									
Completed without help	8	9	8	9	8	4	9	7	6
Needed hint(s)	2	1	2	1	2	6	1	3	4
Failed to complete	0	0	0	0	0	0	0	0	0
Combined, n									
Completed without help	17	18	17	14	15	5	16	12	10
Needed hint(s)	3	1	3	5	5	13	4	8	7
Failed to complete	0	1	0	1	0	2	0	0	3

Phase I—Participant (Patient) Efficiency With Mobile Health Care Environment Technology

Overall, goal setting and manual entry tasks took the least amount of time to complete, and Bluetooth syncing tasks took the most time (Table 2).

Phase I—Participants and System Usability Survey Results

All phase I participants were administered a SUS related to MHCE use. The survey results were very favorable and related to the usability of MHCE. Specifically:

- SUS overall, Mean (SD): 83.8 (SD 14.9)=A+

- Usability sub factor, mean (SD): 86 (SD 13.4)=A+
- Learnability subfactor, mean (SD): 75 (SD 27.5)=A−

Phase I—Results of the Qualitative Research Component (Patient Participants)

All participating participants indicated that the MHCE would "help them manage their diabetes" and give their health care provider a "better report of their health." As expected, patients rated navigation tasks as "less difficult" and peripheral device tasks, for example, syncing, as "more difficult." An overview of usability testing and patient feedback is included in Multimedia Appendix 5. Participants committed fewer errors with basic navigation tasks and more with peripheral devices. Patient participants suggested minor changes regarding the look and function of the app (Textbox 3).

Phase I—Results of the Qualitative Research Component (Clinicians)

Overall, clinicians voiced a sense of optimism for the MHCE system, tempered with recommendations aimed at improving the patient experience, clinician adoption, and use of the data it generated. Clinician participants also suggested changes related to quality and safety (Textbox 4).

Technology Adaptation Following Phase I Research and Before the Launch of Phase II

Patient and clinician suggestions were reviewed and incorporated as adaptations by our technology team within system constraints. A total of 29 specific recommendations were made to the technology team; 86% (25/29) of the recommendations were positively acted on and modifications occurred. Only 4 recommendations could not be supported (Textbox 5).

Table 2. Participant efficiency with technology.

Task	BP[a] goal setting	Glucose goal setting	Weight goal setting	Steps goal setting	BP manual entry	BP sync with device	Glucose manual entry	Weight manual entry	Weight sync with device
Madigan, mean[b] (SD)	55.02 (37.08)	19.83 (7.80)	40.18 (45.57)	61.31 (54.80)	48.00 (31.01)	121.74 (28.09)	32.09 (15.96)	54.09 (38.54)	65.08 (27.18)
Nellis, mean (SD)	51.53 (53.12)	30.29 (18.77)	19.02 (12.67)	19.91 (11.75)	46.44 (32.56)	103.04 (47.05)	30.10 (20.95)	35.37 (23.98)	68.09 (37.23)
Combined sites, mean (SD)	53.18 (45.00)	25.64 (15.50)	29.60 (34.31)	40.61 (44.04)	47.22 (30.97)	111.90 (39.37)	31.09 (18.15)	44.73 (32.69)	66.75 (32.25)

[a]BP: blood pressure.

[b]Mean denotes mean time in seconds.

Textbox 3. Participants' suggestions for Mobile Health Care Environment improvement.

- Suggestion #1: To increase the size of icons as they were difficult to see.
- Suggestion #2: To improve glucose graph responsiveness.
- Suggestion #3: To increase the font size for the Mobile Health Care Environment (MHCE) interface.
- Suggestion #4: To allow past dates for manual entry into the MHCE.

Textbox 4. Clinicians' suggestions related to app safety alert functions.

- Change default blood glucose entry to "Unclassified," forcing patient entry of "after meal," "bedtime," or "fasting."
- Define "after meal" glucose as >120 min post meal.
- Simplify safety alerts on the backend clinician portal.
- Add icons besides safety alerts.
- Add patient target ranges on graphs.

Textbox 5. Recommendations that could not be adopted.

1. Desire for additional colors (limited colors were available).
2. Remove signal and refresh buttons (component of the base system).
3. Add alert icons (limited icon choices available).
4. Automatic syncing of devices (system could not support).

Safety Algorithms Developed and Deployed in Phase II

The research team developed and implemented a series of safety alerts into the MHCE technology. These include thresholds and alerts for high/low blood glucose, high/low blood pressure, and weight loss/gain. Details on safety alerts are identified in a report attached as Multimedia Appendix 6.

Phase II—Participant Randomization, Retention, and Comorbidities

A total of 240 participants met the study eligibility criteria and were randomized (117 to control group and 123 to intervention group). Of the 117 participants randomized to the control group, 5 withdrew before intake, 3 declined participation, and 2 did not meet the inclusion/exclusion criteria. Of the 123 participants randomized to the intervention group, 6 withdrew before intake, 3 declined participation, 1 did not meet the inclusion/exclusion criteria, and 2 did not provide reason. Thus, the final sample size consisted of 229 participants, with 112 randomized to the control group and 117 randomized to the intervention group.

Study retention and participation was high; 93.0% (213/229) of participants remained in the study through 12 months: 99.1% (111/112) in the control group and 86.3% (101/117) in the intervention group. One participant who withdrew from the control group cited a lack of ability to use the physical activity monitor as their primary reason for discontinuation. Of the 16 participants who did not complete the 12-month intervention, 8 cited technical issues with MHCE as their primary reason, 6 listed either a time constraint or relocation as their primary reason, and 2 were lost to follow-up. The specific technical issues cited by the 8 participants withdrawing from the study were not fully known. Anecdotal evidence received by the senior research associates suggested that a few patients were frustrated with system upgrades and challenges with occasional manual synching of devices.

Throughout the 12-month study, 20.6% (21/102) presented new comorbidities in the control group, and 11% (11/96) of subjects presented new comorbidities in the intervention group (χ^2_1=3.2; *P*=.07). On the basis of the covariate-adjusted GEE model, the estimated odds of a new comorbidity throughout the study was 2.27 times higher for the control group than for the intervention group (χ^2_1=3.9; *P*=.047).

Phase II—Baseline Characteristics

Table 3 provides the study baseline demographics and outcomes at baseline. Participants had a mean age of 62.9 (SD 10.3) years, a mean BMI of 32.7 (SD 6.2), and were 61.6% (141/229) male. Mean HbA_{1c} was in the high range (mean 7.5%, SD 1.4%). The mean duration of T2D was 9.8 (SD 7.5) years. The study cohort was well educated; 87.8% (201/229) attended some college and 55.5% (127/229) had a college degree or higher. In this sample, 10% (23/229) of patients had a baseline PAM level of 1 (14/112, 12.5% in control group, 9/117, 7.7% in intervention group), 14.8% (34/229) had a baseline PAM level of 2 (19/112, 17.0% in control group, 15/117, 12.8% in intervention group), 51.5% (118/229) had a baseline PAM level of 3 (55/112, 49.1% in the control group, 63/117, 53.8% in the intervention group), and 23.6% (54/229) had a baseline PAM level of 4 (24/112, 21.4% in control group, 30/117, 25.6% in intervention group). Between group differences were not statistically significant (*P*=.13). The control and intervention groups were well balanced on most baseline characteristics, with the exception of work status (unemployment: 77/112, 68.8% in the control group, 59/117, 50.4% in the intervention group; full-time employment: 27/112, 24.1% in control group, 42/117, 35.9% in intervention group; *P*=.02), baseline low-density lipoprotein (LDL) cholesterol (control group: mean 97.1, SD 39.1; intervention group: mean 80.9 (SD 36.5); *P*=.001). Due to this imbalance, we adjust for these variables in all models.

Table 3. Characteristics of phase II participant demographics and outcomes at baseline.

Variable	Total	Group		*P* value
		Control	Intervention	
Demographics				
Age (years), mean (SD)[a]	62.9 (10.3)	63.2 (9.8)	62.5 (10.8)	.49
Gender (female), n (%)[b]	88 (38.4)	49 (43.8)	39 (33.3)	.11
Race, n (%)[c]				.70
American Indian	2 (0.9)	0 (0.0)	2 (1.7)	
Asian	17 (7.4)	8 (7.1)	9 (7.7)	
Black	33 (14.4)	14 (12.5)	19 (16.2)	
Native Hawaiian	16 (7.0)	7 (6.3)	9 (7.7)	
Unknown/unreported	18 (7.9)	8 (7.1)	10 (8.5)	
White	143 (62.4)	75 (67.0)	68 (58.1)	
Hispanic, n (%)[b]	24 (10.5)	10 (8.9)	14 (12.0)	.45
Education, n (%)[d]				.22
Less than high school	1 (0.4)	1 (0.9)	0.0 (0.0)	
High school graduate	27 (11.8)	15 (13.4)	12 (10.3)	
Some college, no degree	74 (32.3)	35 (31.3)	39 (33.3)	
Associate's degree	41 (17.9)	24 (21.4)	17 (14.5)	
Bachelor's degree	46 (20.1)	20 (17.9)	26 (22.2)	
Master's degree	36 (15.7)	16 (14.3)	20 (17.1)	
Professional degree	4 (1.7)	1 (0.9)	3 (2.6)	
Weight (lbs), mean (SD)[a]	210 (47.1)	209 (50.1)	211 (44.2)	.50
Height (inches), mean (SD)[a]	67.1 (4.0)	67.0 (3.8)	67.2 (4.2)	.44
Duration of type 2 diabetes (years), mean (SD)[a]	9.8 (7.5)	10.0 (7.6)	9.6 (7.4)	.69
Current smoker, n (%)[b]	15 (6.6)	8 (7.1)	7 (6.0)	.72
Hospitalization, n (%)[b]	34 (14.8)	21 (18.8)	13 (11.1)	.10
Emergency room visit, n (%)[b]	87 (38.0)	45 (40.2)	42 (35.9)	.50
Surgeries, n (%)[b]	31 (13.5)	14 (12.5)	17 (14.5)	.65
Work status, n (%)[c]				.02
Unemployed	136 (59.4)	77 (68.8)	59 (50.4)	
Part-time	24 (10.5)	8 (7.1)	16 (13.7)	
Full-time	69 (30.1)	27 (24.1)	42 (35.9)	
Family history of cancer, n (%)[b]	137 (59.8)	72 (64.3)	65 (55.6)	.18
Family history of heart disease, n (%)[b]	132 (57.6)	64 (57.1)	68 (58.1)	.88
Family history of high Blood pressure, n (%)[b]	164 (71.6)	78 (69.6)	86 (73.5)	.52
Family history of high cholesterol, n (%)[b]	111 (48.5)	55 (49.1)	56 (47.9)	.85
Site=Madigan, n (%)[b]	113 (49.3)	51 (45.5)	62 (53.0)	.26
PAM[e], n (%)[d]				.13
Level 1	23 (10.0)	14 (12.5)	9 (7.7)	

Variable	Total	Group		*P* value
		Control	Intervention	
Level 2	34 (14.8)	19 (17.0)	15 (12.8)	
Level 3	118 (51.5)	55 (49.1)	63 (53.8)	
Level 4	54 (23.6)	24 (21.4)	30 (25.6)	
Outcomes				
PAM, mean (SD)[a]	63.2 (12.8)	62.8 (13.5)	63.6 (12.2)	.56
Summary of Diabetes Self-Care Activities score, mean (SD)[a]	37.5 (14.0)	38.0 (14.3)	37.1 (13.8)	.58
Glycated hemoglobin, mean (SD)[a]	7.5 (1.4)	7.6 (1.6)	7.5 (1.3)	.95
BMI, mean (SD)[a]	32.7 (6.2)	32.6 (6.8)	32.8 (5.7)	.77
Waist (inches), mean (SD)[a]	44.1 (5.9)	43.9 (6.4)	44.2 (5.4)	.63
Diastolic BP (mm Hg), mean (SD)[a]	79.5 (9.6)	80.7 (9.4)	78.4 (9.7)	.09
Systolic BP (mm Hg), mean (SD)[a]	133 (17.6)	134 (17.9)	131 (17.3)	.25
High-density lipoprotein cholesterol (mg/dL), mean (SD)[a]	45.2 (13.6)	46.6 (15.4)	43.8 (11.3)	.47
Low-density lipoprotein cholesterol (mg/dL), mean (SD)[a]	88.9 (38.6)	97.1 (39.1)	80.9 (36.5)	.001

[a]Continuous variables compared using Wilcoxon/Kruskal-Wallis test.

[b]Binary variables compared using chi-square test.

[c]Nominal variables compared using Fisher exact test.

[d]Ordinal variables compared using Cochran-Armitage trend test.

[e]PAM: Patient Activation Measure.

Phase II—Primary Outcomes

Within- and between-group comparisons based on GEE are displayed in Table 4. For each intervention group, we assessed changes in the outcome between baseline and month 6 and baseline and month 12. For each group, CIs that exclude 0 indicate a significant change in the outcome between baseline and a given month. *P* values <.05 indicate that this change in the outcome was significantly different between the intervention and control groups. At months 6 and 12, statistically significant improvements were seen in the control group for PAM (month 6: control group improvement=4.81, 95% CI 2.21 to 7.42; month 12: control group improvement=7.49, 95% CI 4.44 to 10.55). That is, improvement in PAM for the control group was 4.81 between baseline and month 6 and 7.49 between baseline and month 12. Although the intervention group showed slight improvement in PAM, these improvements were not statistically significant (month 6: intervention group improvement=1.57, 95% CI −1.34 to 4.48; month 12: intervention group improvement=1.77, 95% CI −1.02 to 4.57). The improvement in the control group was significantly greater than the improvement in the intervention group at month 12 only (month 6: between-group difference=3.24; *P*=.10 and month 12: between-group difference=5.72; *P*=.007).

For each group, estimates (and the corresponding CIs) are given for the *change* in outcome between month 6 and baseline and month 12 and baseline. CIs that exclude 0 indicate statistically significant within-group change between a given month (ie, month 6 or 12) and baseline. The first *P* value column corresponds to testing whether the change in outcomes between month 6 and baseline is different between the 2 groups. The second *P* value column corresponds to testing whether the change in outcomes between month 12 and baseline is different between the 2 groups.

Table 4. Analysis of change from baseline at midpoint and study end.

Outcome	Month 6			Month 12		
	Control, OR[a] (95% CI)	Intervention, OR (95% CI)	*P* value	Control, OR (95% CI)	Intervention, OR (95% CI)	*P* value
Patient Activation Measure score	4.81 (2.21 to 7.42)	1.57 (−1.34 to 4.48)	.10	7.49 (4.44 to 10.55)	1.77 (−1.02 to 4.57)	.007
Summary of Diabetes Self-Care Activities score	7.52 (5.43 to 9.61)	7.52 (5.30 to 9.73)	>.99	6.97 (4.40 to 9.55)	7.52 (5.06 to 9.97)	.77
Glycated hemoglobin	−0.36 (−0.57 to −0.14)	−0.18 (−0.36 to 0.01)	.21	−0.53 (−0.78 to −0.29)	−0.11 (−0.28 to 0.07)	.006
BMI (kg/m^2)	−0.31 (−0.62 to 0.00)	−0.28 (−0.53 to −0.02)	.86	−0.45 (−0.82 to −0.07)	−0.36 (−0.62 to −0.10)	.72
Waist (inches)	−0.57 (−0.96 to −0.19)	−0.76 (−1.09 to −0.44)	.46	−1.46 (−1.95 to −0.98)	−1.49 (−1.90 to −1.07)	.95
Diastolic blood pressure (mm Hg)	−1.83 (−3.81 to 0.15)	−0.58 (−2.30 to 1.14)	.35	−2.18 (−3.98 to −0.38)	−1.72 (−3.22 to −0.22)	.70
Systolic blood pressure (mm Hg)	−2.90 (−6.31 to 0.52)	0.57 (−3.13 to 4.26)	.18	−2.88 (−6.63 to 0.87)	−1.63 (−4.99 to 1.74)	.63
High-density lipoproteincholesterol (mg/dL)	1.22 (−0.12 to 2.57)	1.59 (−0.39 to 3.58)	.76	−1.39 (−2.82 to 0.03)	−0.26 (−2.26 to 1.74)	.37
Low-density lipoprotein cholesterol (mg/dL)	−5.78 (−11.22 to −0.33)	0.12 (−5.69 to 5.93)	.15	−7.14 (−13.43 to −0.85)	4.38 (−2.16 to 10.91)	.01

[a]OR: odds ratio.

Phase II—Secondary Outcomes

At month 6, the control group exhibited statistically significant improvements in HbA_{1c} and LDL cholesterol (as indicated by the CIs in the first column of Table 4). Both intervention and control groups exhibited statistically significant improvements in SDSCA, BMI, and waist size. However, the differences between the intervention and control groups on improvements in these outcomes were not significantly different at month 6 (as indicated by *P* values >.05 in the third column of Table 4). At month 12, statistically significant improvements occurred in SDSCA, BMI, waist size, and diastolic blood pressure in both the intervention and control groups. The control group also exhibited significant improvements in HbA_{1c} and LDL cholesterol at month 12, and these improvements were significantly greater compared with those of the intervention group (HbA_{1c}: between-group difference=−0.42; *P*=.006 and LDL cholesterol: between-group difference=−11.52; *P*=.01).

Phase II—Stratified Analyses by Baseline Patient Activation Measure Level

The intervention is tailored according to baseline PAM level, and therefore, we expect the effect of intervention over time to differ by subjects' baseline PAM level. To test this, a three-way interaction term for the intervention group, time, and baseline PAM level was incorporated into the above GEE models. The results for the change between baseline and end of study (month 12) for each group, stratified by baseline PAM level, are presented in Tables 5 and 6.

Within each baseline PAM level, estimates (and corresponding CIs) are given for the *change* in outcome between month 12 and baseline for both the intervention and control groups. CIs that exclude 0 indicate statistically significant within-group change between month 12 and baseline for subjects in a given PAM level. The *P* value column corresponds to testing whether the change in outcomes is different between the intervention and control groups within a given PAM level.

Among subjects with baseline PAM level 1, both intervention and control groups exhibited significant improvements in PAM score at month 12 (control group improvement=20.38, 95% CI 13.05 to 27.71; intervention group improvement=15.93, 95% CI 8.99 to 22.87); however, between-group differences did not reach statistical significance (*P*=.39). On secondary outcomes, the control group exhibited significant improvements in waist size, diastolic, and systolic blood pressure. The intervention group exhibited significant improvements in SDSCA, HbA_{1c}, BMI, waist size, and HDL cholesterol. Statistically significant between-group differences occurred only in HbA_{1c}, BMI, and HDL cholesterol, indicating that the intervention group exhibited greater improvement in these outcomes compared with the control group (HbA_{1c}: between-group difference=−0.43, *P*=.04; BMI: between-group difference=−1.80, *P*=.01; HDL cholesterol: between-group difference=8.42, *P*<.001).

Table 5. Analysis of change from baseline to study end by baseline Patient Activation Measure level.

Outcome	Baseline PAM[a] level 1			Baseline PAM level 2		
	Control, OR[b] (95% CI)	Intervention, OR (95% CI)	*P* value	Control, OR (95% CI)	Intervention, OR (95% CI)	*P* value
PAM score	20.38 (13.05 to 27.71)	15.93 (8.99 to 22.87)	.39	13.09 (6.26 to 19.92)	7.66 (3.33 to 11.99)	.19
Summary of Diabetes Self-Care Activities score	8.99 (−1.49 to 19.48)	12.78 (3.71 to 21.85)	.59	6.42 (0.49 to 12.36)	11.09 (6.84 to 15.34)	.21
Glycated hemoglobin	−0.09 (−0.33 to 0.16)	−0.52 (−0.87 to −0.18)	.04	−1.04 (−1.88 to −0.20)	−0.09 (−0.59 to 0.42)	.06
BMI (kg/m^2)	0.58 (−0.37 to 1.53)	−1.22 (−2.21 to −0.23)	.01	−0.54 (−1.28 to 0.20)	0.13 (−0.63 to 0.90)	.22
Waist (inches)	−1.50 (−2.73 to −0.27)	−1.98 (−3.32 to −0.64)	.61	−1.42 (−2.36 to −0.47)	−1.90 (−3.30 to −0.51)	.57
Diastolic blood pressure (mm Hg)	−4.00 (−7.86 to −0.14)	−3.89 (−10.79 to 3.01)	.98	−3.22 (−7.38 to 0.95)	−0.50 (−4.18 to 3.18)	.34
Systolic blood pressure (mm Hg)	−9.93 (−17.18 to −2.68)	−8.33 (−17.17 to 0.51)	.78	5.86 (−3.95 to 15.67)	−2.42 (−10.76 to 5.92)	.21
High-density lipoprotein cholesterol (mg/dL)	−4.86 (−8.16 to −1.56)	3.56 (0.50 to 6.61)	<.001	1.40 (−3.22 to 6.02)	−2.76 (−6.30 to 0.77)	.16
Low-density lipoprotein cholesterol (mg/dL)	−10.21 (−25.74 to 5.31)	9.56 (−10.02 to 29.13)	.12	−5.62 (−24.94 to 13.70)	10.55 (−12.21 to 33.30)	.29

[a]PAM: Patient Activation Measure.

[b]OR: odds ratio.

Table 6. Analysis of change from baseline to study end baseline Patient Activation Measure level (continued).

Outcome	Baseline PAM[a] level 3			Baseline PAM level 4		
	Control, OR[b] (95% CI)	Intervention, OR (95% CI)	*P* value	Control, OR (95% CI)	Intervention, OR (95% CI)	*P* value
PAM score	6.29 (2.24 to 10.35)	4.26 (0.88 to 7.64)	.45	−1.44 (−7.39 to 4.50)	−10.49 (−16.54 to −4.44)	.04
Summary of Diabetes Self-Care Activities score	6.49 (2.75 to 10.23)	7.58 (4.44 to 10.73)	.66	7.02 (3.15 to 10.89)	3.95 (−1.76 to 9.66)	.38
Glycated hemoglobin	−0.49 (−0.83 to −0.16)	−0.04 (−0.25 to 0.16)	.02	−0.57 (−1.11 to −0.03)	−0.09 (−0.55 to 0.36)	.18
BMI (kg/m^2)	−0.43 (−0.86 to 0.01)	−0.42 (−0.76 to −0.07)	.97	−1.25 (−2.47 to −0.04)	−0.09 (−0.53 to 0.35)	.08
Waist (inches)	−1.51 (−2.18 to −0.83)	−1.56 (−2.08 to −1.03)	.91	−1.37 (−2.63 to −0.10)	−0.90 (−1.71 to −0.08)	.54
Diastolic blood pressure (mm Hg)	−2.55 (−5.22 to 0.11)	−1.23 (−3.24 to 0.79)	.44	−0.19 (−4.24 to 3.86)	−2.22 (−5.41 to 0.96)	.44
Systolic blood pressure (mm Hg)	−4.33 (−9.20 to 0.53)	−1.58 (−6.34 to 3.19)	.43	−2.78 (−12.44 to 6.89)	1.69 (−5.17 to 8.56)	.46
High-density lipoprotein cholesterol (mg/dL)	−0.93 (−2.84 to 0.99)	0.12 (−2.56 to 2.81)	.53	−2.44 (−4.91 to 0.02)	−1.03 (−6.10 to 4.05)	.62
Low-density lipoprotein cholesterol (mg/dL)	−3.99 (−12.36 to 4.37)	3.32 (−3.62 to 10.26)	.19	−13.38 (−27.02 to 0.25)	−0.19 (−17.09 to 16.70)	.23

[a]PAM: Patient Activation Measure.

[b]OR: odds ratio.

Among subjects with baseline PAM levels 2 and 3, both the intervention and control groups exhibited significant improvements in PAM, SDSCA, and waist size. However, between-group differences were not statistically significant. The control group also exhibited significant improvements in HbA_{1c}; between-group differences were significant among those with baseline PAM level 3 only (between-group difference=−0.45; *P*=.02). The intervention group also exhibited significant improvement in BMI among subjects with PAM level 3, but the between-group differences did not reach statistical significance.

Among subjects with baseline PAM level 4, the intervention group exhibited a significant decrease in PAM (intervention group change=−10.49; 95% CI −16.54 to −4.44), whereas control subjects exhibited no significant change (control group change=−1.44; 95% CI −7.39 to 4.50). This difference in change

was statistically significant (between-group difference=−9.05; *P*=.04). The control group exhibited significant improvements in SDSCA, HbA_{1c}, and BMI. Both groups exhibited significant improvements in waist size. However, between-group differences did not reach statistical significance in these outcomes.

Phase II—Participant Engagement and Usability

The percentage of subjects who engaged with MHCE at least 50% of the days in each time period was 60.7% (68/112) for months 0-3, 57.4% (62/108) for months 3-6, 49.5% (51/103) for months 6-9, and 43% (42/98) for months 9-12. The percentage of subjects who engaged with MHCE at least 80% of the days in each time period was 41.1% (46/112) for months 0-3, 41.7% (45/108) for months 3-6, 14.6% (15/103) for months 6-9, and 26% (25/98) for months 9-12. Strong adherence, defined engagement with MHCE on at least 80% of days in each time period, was associated with improvement in all outcomes. However, only the association with HbA_{1c} reached statistical significance (mean improvement=0.35; *P*=.049).

The mean SUS score for the MHCE group was 75.6 (SD 17.3) at month 6 and 76.8 (SD 14.8) at month 12. Strong adherence was associated with a higher SUS score (average increase=5.01; *P*=.04). As SUS was not recorded for the control group, group differences could not be compared.

Discussion

Principal Findings

The primary goal of this large feasibility study was to enhance the PAM level and improve self-management of T2D care using the DoD's interactive and tailored MHCE system for T2D in a PCMH setting. To this end, we hypothesized that a user-centered design process would successfully shape the development of the MHCE for use in the self-management of T2D care. The research team engaged both patients and clinicians, observing patient and clinician overall satisfaction with the MHCE, high SUS (A+ and A−) scores, and 29 offered recommendations for system improvement; 86% (25/29) of these recommendations resulted in intervention modifications before feasibility testing. The methods and findings from the study's phase I research align well with the aging barriers addressed by Wildenbos et al [61]. These include cognitive, physical ability, perception, and motivational barriers.

When comparing this study with comparable mHealth studies, there were differences in participant demographics. Specifically, our research participants tended to be older (age: mean 62.9 years, SD 10.3) than participants in other studies [36,38,62-69] and had a larger BMI (32.7 kg/m^2, SD 6.2 kg/m^2) at baseline [62,64-66,68]. This study engaged a substantial percentage, 59.4% (136/229), of unemployed participants; most of them were military retirees. Finally, our research participants were members of a PCMH site within one of two military treatment facilities (ie, military run hospitals) and as such typically received health care services with no (or minimal) out-of-pocket costs. A descriptive primer on this federal health system has been published [70].

Overall retention was high in this study, 93.0% (213/229) for the 12-month study, which is consistent with other mHealth projects [66] and demonstrates the feasibility of conducting a PCMH intervention among T2D patients. The 12-month duration of this study was longer than loosely comparable studies, which generally targeted a 9-month [36,63], 6-month [64,65], or 3-month duration [38,62].

The team hypothesized that the use of the MHCE would increase the PAM and self-care activities of patients with T2D and that patients who engage at a higher rate with the MHCE would realize improvement in clinical measures. In our study, both intervention and intervention-lite (control) groups showed significant improvement from baseline to 12 months in SDSCA (measuring self-care activities) as well as in BMI, waist size, and diastolic blood pressure control. A somewhat surprising finding in the study is that only the control group exhibited significant improvement in the PAM score. Furthermore, improvements in the control group exceeded improvements in the intervention group for most clinical outcomes.

However, a closer analysis of the data reveals a complex yet intriguing pattern of how mHealth intervention effectiveness can vary by the baseline level of patient activation. With the exception of patients with the highest level of activation (PAM level 4), all intervention group patients exhibited significant improvements in the PAM score. Furthermore, the intervention appears to be superior to control for patients with the lowest level of activation (PAM level 1). In this class of patients, the intervention group showed significantly greater improvements at the end of the study in HbA_{1c}, BMI, and HDL cholesterol levels compared with the control group (these were the only significant differences between the groups). This finding may suggest that in a resource-limited health system environment, a targeted investment in mHealth to support self-management of T2D for PAM level 1 patients may yield optimal results compared with patients who are more activated. With the more activated patient, demonstrating higher PAM scores, a less costly peripheral device (eg, glucometer, activity monitor, and scale) may by itself enhance self-care activities.

Our finding that the intervention group exhibited significant improvements over the control group in several cardiovascular outcomes in T2D patients with baseline PAM level 1 is consistent with previous findings that text messaging is effective in helping patients with T2D achieve better glycemic control [16,36,71]. Although tailored messaging did not generate superior results compared with the control group for patients with higher PAM levels, it does not suggest that the PAM-aligned tailored behavioral messaging was faulty or should be discarded. In our study, the DoD technology partners secured the tailored behavioral messaging embedded within the MHCE under an indicator labeled as "health tips." However, because of privacy concerns, patients could only see the message after logging into the MHCE mobile app through its security layers to view the message. The multistep process of accessing the messages was indirect and could have been overlooked on the patients' personal cell phone. Future efforts should consider redesigning the MHCE (or mHealth platforms) to accommodate nonpersonal health information (PHI) and tailored health

message delivery to the patient in his/her preferred communication process. Most of this will likely be via mobile app push notifications and/or text messaging. Several published studies have successfully leveraged tailored health messaging (without PHI), delivered in a direct delivery process via text messaging to participants [16,36,69]; one study was very short term in nature [69].

Although SUSs remained constant throughout the study and were associated with increased patient engagement, overall patient engagement with the MHCE system decreased throughout the study period. Decreased engagement may be partly explained by frustration with system upgrades and synching of devices expressed by some patients, along with the multistep process of MHCE discussed above. However, high levels of initial patient engagement in other mHealth projects that drop off after 6 to 12 months have been observed [16,63,72]. This might suggest that systems such as the MHCE may be optimally deployed in self-care of T2D for periods up to about 6 months; lengthy engagement (>6 months) may not generate the desired results.

Limitations

Our study design is not immune from potential threats to validity. Both the intervention and control groups were issued the same clinical outcome measuring devices and completed the biweekly SDSCA survey. However, we do not have information on whether patients in the control group did the synchronization of their measuring devices to their mobile phones, tablet computers, or desktop/laptop computers for viewing and tracking purposes. Therefore, we cannot rule out the possibility that a substantial proportion of the control group had access to visualized tracking information similar to the intervention group. For these patients, improvement in these outcomes may be explained by the same mechanisms whereby self-weighing is linked with weight loss [73]. Another limitation of the study is that food intake and daily activities outside the research sites are uncontrolled due to the nature of this type of field experiment.

Finally, although our study included user-center design testing (including solicitation of user design preferences) and ultimately generated a high retention rate for the 12-month study, patient engagement with MHCE decreased throughout the course of the phase II study. This trend was also seen in previous mHealth projects, and future studies should therefore consider incorporating follow-up participant assessments of continued engagement. If implemented following phase I, this may help the understanding of discontinued engagement and provide opportunities for improvement.

Acknowledgments

Research was extramurally funded by the US Army Medical Research Acquisition Activity contract #W81XWH-15-C-0070. The funder did not influence the design of the study or strategies related to its collection, analysis, or interpretation. The views expressed are those of the authors and do not reflect the official policy of the US Department of the Army, the Department of the Air Force, the Department of Defense, or the US Federal Government.

Authors' Contributions

RG initially conceptualized the study, and both RG and LR wrote the initial draft of the manuscript. PC, RG, KE, MH, KC, and JM assisted in the setting of the narrative, enrollment strategies, and institutional review board–related issues. The technical aspects of the intervention and MHCE description were authored by JL and RG. The biostatistics and data analysis strategies were developed by LC, LS, and LR. The patient activation activities in the study were coordinated by MH, KC, JM, AS, KP, and KE. The user-centered design and qualitative component were researched and authored by JW, SG, and KE. The outcome measures component was conceptualized and authored by PC, KT, LS, LZ, JW, LC, MH, and KE. The clinical components, including the inclusion and exclusion criteria, were developed by PC, JM, MH, KE, and RG. The Military Health System review and components were authored by JM, MH, KE, and RG. Active site Principal Investigators in the study were PC and RM. Acquisition and formatting of data for analysis were conducted by KT, LC, MA, and EH. Statistical analyses were conducted by LR, KT, BL, and LS. Interpretation of the data was performed by LR, RG, LS, and KT. All authors read, contributed to, critically reviewed, and approved the final manuscript.

Conflicts of Interest

JL and AS are federal employees charged with oversight of the DoD’s MHCE mHealth system used in the study. None of the other authors report any conflicts of interest.

Multimedia Appendix 1
Annotated visual presentation of patient screenshots and workflow.
[PPTX File , 5193 KB - jmir_v22i5e17968_app1.pptx]

Multimedia Appendix 2
Annotated visual presentation of clinician access.
[PPTX File , 4989 KB - jmir_v22i5e17968_app2.pptx]

Multimedia Appendix 3
Phase I data collection instrument.
[PDF File (Adobe PDF File), 195 KB - jmir_v22i5e17968_app3.pdf]

Multimedia Appendix 4
Clinician focus group - facilitator's guide.
[PDF File (Adobe PDF File), 125 KB - jmir_v22i5e17968_app4.pdf]

Multimedia Appendix 5
Mobile Health Care Environment modifications report to funder.
[PDF File (Adobe PDF File), 585 KB - jmir_v22i5e17968_app5.pdf]

Multimedia Appendix 6
Standards, alerts and safety algorithms report.
[PDF File (Adobe PDF File), 211 KB - jmir_v22i5e17968_app6.pdf]

References

1. Sudore RL, Karter AJ, Huang ES, Moffet HH, Laiteerapong N, Schenker Y, et al. Symptom burden of adults with type 2 diabetes across the disease course: diabetes & aging study. J Gen Intern Med 2012 Dec;27(12):1674-1681 [FREE Full text] [doi: 10.1007/s11606-012-2132-3] [Medline: 22854982]
2. Vigersky RA. An overview of management issues in adult patients with type 2 diabetes mellitus. J Diabetes Sci Technol 2011 Mar 1;5(2):245-250 [FREE Full text] [doi: 10.1177/193229681100500207] [Medline: 21527089]
3. Centers for Disease Control and Prevention. 2018. Division of Diabetes Translation At A Glance URL: https://www.cdc.gov/chronicdisease/resources/publications/aag/diabetes.htm [accessed 2019-11-01]
4. Scollan-Koliopoulos M, Bleich D, Rapp KJ, Wong P, Hofmann CJ, Raghuwanshi M. Health-related quality of life, disease severity, and anticipated trajectory of diabetes. Diabetes Educ 2013;39(1):83-91. [doi: 10.1177/0145721712467697] [Medline: 23174664]
5. Wang Y, Xue H, Huang Y, Huang L, Zhang D. A systematic review of application and effectiveness of mhealth interventions for obesity and diabetes treatment and self-management. Adv Nutr 2017 May;8(3):449-462 [FREE Full text] [doi: 10.3945/an.116.014100] [Medline: 28507010]
6. Fatehi F, Gray LC, Russell AW. Mobile health (mhealth) for diabetes care: opportunities and challenges. Diabetes Technol Ther 2017 Jan;19(1):1-3. [doi: 10.1089/dia.2016.0430] [Medline: 28099051]
7. Snider JT, Sullivan J, van Eijndhoven E, Hansen MK, Bellosillo N, Neslusan C, et al. Lifetime benefits of early detection and treatment of diabetic kidney disease. PLoS One 2019;14(5):e0217487 [FREE Full text] [doi: 10.1371/journal.pone.0217487] [Medline: 31150444]
8. Avogaro A, Fadini GP. Microvascular complications in diabetes: a growing concern for cardiologists. Int J Cardiol 2019 Sep 15;291:29-35. [doi: 10.1016/j.ijcard.2019.02.030] [Medline: 30833106]
9. American Diabetes Association. Economic costs of diabetes in the US In 2017. Diabetes Care 2018 May;41(5):917-928 [FREE Full text] [doi: 10.2337/dci18-0007] [Medline: 29567642]
10. Arora S, Peters AL, Agy C, Menchine M. A mobile health intervention for inner city patients with poorly controlled diabetes: proof-of-concept of the TExT-MED program. Diabetes Technol Ther 2012 Jun;14(6):492-496. [doi: 10.1089/dia.2011.0252] [Medline: 22524591]
11. Mohsin AH, Zaidan AA, Zaidan BB, Albahri OS, Albahri AS, Alsalem MA, et al. Based medical systems for patient's authentication: towards a new verification secure framework using CIA standard. J Med Syst 2019 May 22;43(7):192. [doi: 10.1007/s10916-019-1264-y] [Medline: 31115768]
12. Capozza K, Woolsey S, Georgsson M, Black J, Bello N, Lence C, et al. Going mobile with diabetes support: a randomized study of a text message-based personalized behavioral intervention for type 2 diabetes self-care. Diabetes Spectr 2015 May;28(2):83-91 [FREE Full text] [doi: 10.2337/diaspect.28.2.83] [Medline: 25987806]
13. Sahin C, Courtney KL, Naylor PJ, Rhodes RE. Tailored mobile text messaging interventions targeting type 2 diabetes self-management: a systematic review and a meta-analysis. Digit Health 2019;5:2055207619845279 [FREE Full text] [doi: 10.1177/2055207619845279] [Medline: 31041110]
14. Borrelli B, Henshaw M, Endrighi R, Adams WG, Heeren T, Rosen RK, et al. An interactive parent-targeted text messaging intervention to improve oral health in children attending urban pediatric clinics: feasibility randomized controlled trial. JMIR Mhealth Uhealth 2019 Nov 11;7(11):e14247 [FREE Full text] [doi: 10.2196/14247] [Medline: 31710306]
15. Kruse C, Betancourt J, Ortiz S, Luna SM, Bamrah IK, Segovia N. Barriers to the use of mobile health in improving health outcomes in developing countries: systematic review. J Med Internet Res 2019 Oct 9;21(10):e13263 [FREE Full text] [doi: 10.2196/13263] [Medline: 31593543]

16. McGill DE, Volkening LK, Butler DA, Wasserman RM, Anderson BJ, Laffel LM. Text-message responsiveness to blood glucose monitoring reminders is associated with HbA benefit in teenagers with type 1 diabetes. Diabet Med 2019 May;36(5):600-605. [doi: 10.1111/dme.13929] [Medline: 30734361]
17. Chung S, Panattoni L, Chi J, Palaniappan L. Can secure patient-provider messaging improve diabetes care? Diabetes Care 2017 Oct;40(10):1342-1348. [doi: 10.2337/dc17-0140] [Medline: 28807977]
18. Sun R, Korytkowski MT, Sereika SM, Saul MI, Li D, Burke LE. Patient portal use in diabetes management: literature review. JMIR Diabetes 2018 Nov 6;3(4):e11199 [FREE Full text] [doi: 10.2196/11199] [Medline: 30401665]
19. Gram IT, Larbi D, Wangberg SC. Comparing the efficacy of an identical, tailored smoking cessation intervention delivered by mobile text messaging versus email: randomized controlled trial. JMIR Mhealth Uhealth 2019 Sep 27;7(9):e12137 [FREE Full text] [doi: 10.2196/12137] [Medline: 31573935]
20. Ukoha EP, Feinglass J, Yee LM. Disparities in electronic patient portal use in prenatal care: retrospective cohort study. J Med Internet Res 2019 Sep 23;21(9):e14445 [FREE Full text] [doi: 10.2196/14445] [Medline: 31586367]
21. Gimbel R, Shi L, Williams JE, Dye CJ, Chen L, Crawford P, et al. Enhancing mhealth technology in the patient-centered medical home environment to activate patients with type 2 diabetes: a multisite feasibility study protocol. JMIR Res Protoc 2017 Mar 6;6(3):e38 [FREE Full text] [doi: 10.2196/resprot.6993] [Medline: 28264792]
22. Greene J, Hibbard JH, Sacks R, Overton V, Parrotta CD. When patient activation levels change, health outcomes and costs change, too. Health Aff (Millwood) 2015 Mar;34(3):431-437. [doi: 10.1377/hlthaff.2014.0452] [Medline: 25732493]
23. Mattingly TJ, Tom SE, Stuart B, Onukwugha E. Examining patient-provider relationship (PPR) quality and patient activation in the medicare population. Aging Clin Exp Res 2017 Jun;29(3):543-548. [doi: 10.1007/s40520-016-0600-z] [Medline: 27324691]
24. Cronin R, Dorner T, Utrankar A, Allen W, Rodeghier M, Kassim A, et al. Increased patient activation is associated with fewer emergency room visits and hospitalizations for pain in adults with sickle cell disease. Pain Med 2019 Aug 1;20(8):1464-1471. [doi: 10.1093/pm/pny194] [Medline: 30329108]
25. Queenan C, Cameron K, Snell A, Smalley J, Joglekar N. Patient heal thyself: reducing hospital readmissions with technology‐enabled continuity of care and patient activation. Prod Oper Manag 2019 Aug 11;28(11):2841-2853. [doi: 10.1111/poms.13080]
26. Ammenwerth E, Hoerbst A, Lannig S, Mueller G, Siebert U, Schnell-Inderst P. Effects of adult patient portals on patient empowerment and health-related outcomes: a systematic review. Stud Health Technol Inform 2019 Aug 21;264:1106-1110. [doi: 10.3233/SHTI190397] [Medline: 31438096]
27. Schnock KO, Snyder JE, Fuller TE, Duckworth M, Grant M, Yoon C, et al. Acute care patient portal intervention: portal use and patient activation. J Med Internet Res 2019 Jul 18;21(7):e13336 [FREE Full text] [doi: 10.2196/13336] [Medline: 31322123]
28. Denneson L, Pisciotta M, Hooker E, Trevino A, Dobscha S. Impacts of a web-based educational program for veterans who read their mental health notes online. J Am Med Inform Assoc 2019 Jan 1;26(1):3-8. [doi: 10.1093/jamia/ocy134] [Medline: 30445648]
29. Ammenwerth E, Schnell-Inderst P, Hoerbst A. Patient empowerment by electronic health records: first results of a systematic review on the benefit of patient portals. Stud Health Technol Inform 2011;165:63-67. [Medline: 21685587]
30. Irizarry T, DeVito Dabbs A, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015 Jun 23;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
31. Riippa I, Linna M, Rönkkö I. The effect of a patient portal with electronic messaging on patient activation among chronically ill patients: controlled before-and-after study. J Med Internet Res 2014 Nov 19;16(11):e257 [FREE Full text] [doi: 10.2196/jmir.3462] [Medline: 25413368]
32. Apter AJ, Bryant-Stephens T, Perez L, Morales KH, Howell JT, Mullen AN, et al. Patient portal usage and outcomes among adult patients with uncontrolled asthma. J Allergy Clin Immunol Pract 2020 Mar;8(3):965-70.e4. [doi: 10.1016/j.jaip.2019.09.034] [Medline: 31622684]
33. Otte-Trojel T, de Bont A, Rundall TG, van de Klundert J. How outcomes are achieved through patient portals: a realist review. J Am Med Inform Assoc 2014;21(4):751-757 [FREE Full text] [doi: 10.1136/amiajnl-2013-002501] [Medline: 24503882]
34. Wagner PJ, Dias J, Howard S, Kintziger KW, Hudson MF, Seol Y, et al. Personal health records and hypertension control: a randomized trial. J Am Med Inform Assoc 2012;19(4):626-634 [FREE Full text] [doi: 10.1136/amiajnl-2011-000349] [Medline: 22234404]
35. Schuit AS, Holtmaat K, Hooghiemstra N, Jansen F, Lissenberg-Witte BI, Coupé VM, et al. Efficacy and cost-utility of the ehealth application 'Oncokompas', supporting patients with incurable cancer in finding optimal palliative care, tailored to their quality of life and personal preferences: a study protocol of a randomized controlled trial. BMC Palliat Care 2019 Oct 23;18(1):85 [FREE Full text] [doi: 10.1186/s12904-019-0468-8] [Medline: 31647011]
36. Dobson R, Whittaker R, Jiang Y, Maddison R, Shepherd M, McNamara C, et al. Effectiveness of text message based, diabetes self management support programme (SMS4BG): two arm, parallel randomised controlled trial. Br Med J 2018 May 17;361:k1959 [FREE Full text] [doi: 10.1136/bmj.k1959] [Medline: 29773539]

37. Almutairi N, Hosseinzadeh H, Gopaldasani V. The effectiveness of patient activation intervention on type 2 diabetes mellitus glycemic control and self-management behaviors: a systematic review of RCTs. Prim Care Diabetes 2020 Feb;14(1):12-20. [doi: 10.1016/j.pcd.2019.08.009] [Medline: 31543458]
38. Su D, Michaud TL, Estabrooks P, Schwab RJ, Eiland LA, Hansen G, et al. Diabetes management through remote patient monitoring: the importance of patient activation and engagement with the technology. Telemed J E Health 2019 Oct;25(10):952-959. [doi: 10.1089/tmj.2018.0205] [Medline: 30372366]
39. Kim JY, Wineinger NE, Steinhubl SR. The influence of wireless self-monitoring program on the relationship between patient activation and health behaviors, medication adherence, and blood pressure levels in hypertensive patients: a substudy of a randomized controlled trial. J Med Internet Res 2016 Jun 22;18(6):e116 [FREE Full text] [doi: 10.2196/jmir.5429] [Medline: 27334418]
40. Zimbudzi E, Lo C, Ranasinha S, Kerr PG, Polkinghorne KR, Teede H, et al. The association between patient activation and self-care practices: a cross-sectional study of an Australian population with comorbid diabetes and chronic kidney disease. Health Expect 2017 Dec;20(6):1375-1384 [FREE Full text] [doi: 10.1111/hex.12577] [Medline: 28675539]
41. McCabe PJ, Stuart-Mullen LG, McLeod CJ, O Byrne T, Schmidt MM, Branda ME, et al. Patient activation for self-management is associated with health status in patients with atrial fibrillation. Patient Prefer Adherence 2018;12:1907-1916 [FREE Full text] [doi: 10.2147/PPA.S172970] [Medline: 30288031]
42. Greene J, Hibbard JH. Why does patient activation matter? An examination of the relationships between patient activation and health-related outcomes. J Gen Intern Med 2012 May;27(5):520-526 [FREE Full text] [doi: 10.1007/s11606-011-1931-2] [Medline: 22127797]
43. Hibbard J, Stockard J, Mahoney E, Tusler M. Development of the patient activation measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004 Aug;39(4 Pt 1):1005-1026 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]
44. 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]
45. Sauro J, Dumas J. Comparison of Three One-question, Post-task Usability Questionnaires. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2009 Presented at: CHI'09; April 4-9, 2009; Boston, MA. [doi: 10.1145/1518701.1518946]
46. Sauro J, Lewis J. Quantifying the User Experience: Practical Statistics for User Research. Waltham, MA: Morgan Kaufmann; 2012.
47. Kidd PS, Parshall MB. Getting the focus and the group: enhancing analytical rigor in focus group research. Qual Health Res 2000 May;10(3):293-308. [doi: 10.1177/104973200129118453] [Medline: 10947477]
48. Thorn S. Data analysis in qualitative research. Evid Based Nurs 2000;3:68-70. [doi: 10.1136/ebn.3.3.68]
49. Thomas DR. A general inductive approach for analyzing qualitative evaluation data. Am J Eval 2016 Jun 30;27(2):237-246. [doi: 10.1177/1098214005283748]
50. Harvey L, Fowles JB, Xi M, Terry P. When activation changes, what else changes? the relationship between change in patient activation measure (PAM) and employees' health status and health behaviors. Patient Educ Couns 2012 Aug;88(2):338-343. [doi: 10.1016/j.pec.2012.02.005] [Medline: 22459636]
51. Mosen DM, Schmittdiel J, Hibbard J, Sobel D, Remmers C, Bellows J. Is patient activation associated with outcomes of care for adults with chronic conditions? J Ambul Care Manage 2007;30(1):21-29. [doi: 10.1097/00004479-200701000-00005] [Medline: 17170635]
52. Rask KJ, Ziemer DC, Kohler SA, Hawley JN, Arinde FJ, Barnes CS. Patient activation is associated with healthy behaviors and ease in managing diabetes in an indigent population. Diabetes Educ 2009;35(4):622-630. [doi: 10.1177/0145721709335004] [Medline: 19419972]
53. Rogvi S, Tapager I, Almdal T, Schiøtz ML, Willaing I. Patient factors and glycaemic control-associations and explanatory power. Diabet Med 2012 Oct;29(10):e382-e389. [doi: 10.1111/j.1464-5491.2012.03703.x] [Medline: 22540962]
54. Toobert DJ, Hampson SE, Glasgow RE. The summary of diabetes self-care activities measure: results from 7 studies and a revised scale. Diabetes Care 2000 Jul;23(7):943-950 [FREE Full text] [doi: 10.2337/diacare.23.7.943] [Medline: 10895844]
55. Gonzalez J, Safren S, Delahanty L, Cagliero E, Wexler D, Meigs J, et al. Symptoms of depression prospectively predict poorer self-care in patients with type 2 diabetes. Diabet Med 2008 Sep;25(9):1102-1107 [FREE Full text] [doi: 10.1111/j.1464-5491.2008.02535.x] [Medline: 19183315]
56. Watkins YJ, Quinn LT, Ruggiero L, Quinn MT, Choi Y. Spiritual and religious beliefs and practices and social support's relationship to diabetes self-care activities in African Americans. Diabetes Educ 2013;39(2):231-239 [FREE Full text] [doi: 10.1177/0145721713475843] [Medline: 23411653]
57. Primožič S, Tavčar R, Avbelj M, Dernovšek MZ, Oblak MR. Specific cognitive abilities are associated with diabetes self-management behavior among patients with type 2 diabetes. Diabetes Res Clin Pract 2012 Jan;95(1):48-54. [doi: 10.1016/j.diabres.2011.09.004] [Medline: 21963107]

58. Cohen H, Shmukler C, Ullman R, Rivera C, Walker E. Measurements of medication adherence in diabetic patients with poorly controlled HbA(1c). Diabet Med 2010 Feb;27(2):210-216 [FREE Full text] [doi: 10.1111/j.1464-5491.2009.02898.x] [Medline: 20546266]
59. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum-Comput Int 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
60. Diggle P, Heagerty P, Liang K, Zeger S. Analysis of Longitudinal Data. Second Edition. Oxford, UK: Oxford University Press; 2002.
61. Wildenbos GA, Peute L, Jaspers M. Aging barriers influencing mobile health usability for older adults: a literature based framework (MOLD-US). Int J Med Inform 2018 Jun;114:66-75. [doi: 10.1016/j.ijmedinf.2018.03.012] [Medline: 29673606]
62. Bender MS, Cooper BA, Park LG, Padash S, Arai S. A feasible and efficacious mobile-phone based lifestyle intervention for Filipino Americans with type 2 diabetes: randomized controlled trial. JMIR Diabetes 2017 Dec 12;2(2):e30 [FREE Full text] [doi: 10.2196/diabetes.8156] [Medline: 30291068]
63. Young H, Miyamoto S, Dharmar M, Tang-Feldman Y. Nurse coaching and mobile health compared with usual care to improve diabetes self-efficacy for persons with type 2 diabetes: randomized controlled trial. JMIR Mhealth Uhealth 2020 Mar 2;8(3):e16665 [FREE Full text] [doi: 10.2196/16665] [Medline: 32130184]
64. Block G, Azar KM, Romanelli RJ, Block TJ, Hopkins D, Carpenter HA, et al. Diabetes prevention and weight loss with a fully automated behavioral intervention by email, web, and mobile phone: a randomized controlled trial among persons with prediabetes. J Med Internet Res 2015 Oct 23;17(10):e240 [FREE Full text] [doi: 10.2196/jmir.4897] [Medline: 26499966]
65. Koot D, Goh PS, Lim RS, Tian Y, Yau TY, Tan NC, et al. A mobile lifestyle management program (GlycoLeap) for people with type 2 diabetes: single-arm feasibility study. JMIR Mhealth Uhealth 2019 May 24;7(5):e12965 [FREE Full text] [doi: 10.2196/12965] [Medline: 31127720]
66. Everett E, Kane B, Yoo A, Dobs A, Mathioudakis N. A novel approach for fully automated, personalized health coaching for adults with prediabetes: pilot clinical trial. J Med Internet Res 2018 Feb 27;20(2):e72 [FREE Full text] [doi: 10.2196/jmir.9723] [Medline: 29487046]
67. Bell AM, Fonda SJ, Walker MS, Schmidt V, Vigersky RA. Mobile phone-based video messages for diabetes self-care support. J Diabetes Sci Technol 2012 Mar 1;6(2):310-319 [FREE Full text] [doi: 10.1177/193229681200600214] [Medline: 22538140]
68. Spring B, Schneider K, McFadden HG, Vaughn J, Kozak AT, Smith M, et al. Multiple behavior changes in diet and activity: a randomized controlled trial using mobile technology. Arch Intern Med 2012 May 28;172(10):789-796 [FREE Full text] [doi: 10.1001/archinternmed.2012.1044] [Medline: 22636824]
69. Nelson LA, Mayberry LS, Wallston K, Kripalani S, Bergner EM, Osborn CY. Development and usability of REACH: a tailored theory-based text messaging intervention for disadvantaged adults with type 2 diabetes. JMIR Hum Factors 2016 Sep 8;3(2):e23 [FREE Full text] [doi: 10.2196/humanfactors.6029] [Medline: 27609738]
70. Gimbel RW, Pangaro L, Barbour G. America's 'undiscovered' laboratory for health services research. Med Care 2010 Aug;48(8):751-756. [doi: 10.1097/MLR.0b013e3181e35be8] [Medline: 20613659]
71. Saffari M, Ghanizadeh G, Koenig HG. Health education via mobile text messaging for glycemic control in adults with type 2 diabetes: a systematic review and meta-analysis. Prim Care Diabetes 2014 Dec;8(4):275-285. [doi: 10.1016/j.pcd.2014.03.004] [Medline: 24793589]
72. Sepah SC, Jiang L, Ellis RJ, McDermott K, Peters AL. Engagement and outcomes in a digital diabetes prevention program: 3-year update. BMJ Open Diabetes Res Care 2017;5(1):e000422 [FREE Full text] [doi: 10.1136/bmjdrc-2017-000422] [Medline: 28948027]
73. Zheng Y, Burke LE, Danford CA, Ewing LJ, Terry MA, Sereika SM. Patterns of self-weighing behavior and weight change in a weight loss trial. Int J Obes (Lond) 2016 Sep;40(9):1392-1396. [doi: 10.1038/ijo.2016.68] [Medline: 27113642]

Abbreviations

DoD: US Department of Defense
GEE: generalized estimating equations
HbA_{1c}: glycated hemoglobin
LDL: low-density lipoprotein
MHCE: Mobile Health Care Environment
mHealth: mobile health
PAM: Patient Activation Measure
PCMH: patient-centered medical home
PHI: personal health information
PHR: personal health record
SDSCA: Summary of Diabetes Self-Care Activities
SUS: System Usability Scale

T2D: type 2 diabetes

Edited by G Eysenbach; submitted 24.01.20; peer-reviewed by S Jalil, N Lindström, B Eapen, H Hah, I Mircheva; comments to author 14.02.20; revised version received 21.03.20; accepted 12.04.20; published 26.05.20.

Please cite as:
Gimbel RW, Rennert LM, Crawford P, Little JR, Truong K, Williams JE, Griffin SF, Shi L, Chen L, Zhang L, Moss JB, Marshall RC, Edwards KW, Crawford KJ, Hing M, Schmeltz A, Lumsden B, Ashby M, Haas E, Palazzo K
Enhancing Patient Activation and Self-Management Activities in Patients With Type 2 Diabetes Using the US Department of Defense Mobile Health Care Environment: Feasibility Study
J Med Internet Res 2020;22(5):e17968
URL: http://www.jmir.org/2020/5/e17968/
doi:10.2196/17968
PMID:32329438

Original Paper

Digital Biomarkers of Social Anxiety Severity: Digital Phenotyping Using Passive Smartphone Sensors

Nicholas C Jacobson[1], PhD; Berta Summers[2], PhD; Sabine Wilhelm[2], PhD

[1]Center for Technology and Behavioral Health, Geisel School of Medicine, Dartmouth College, Lebanon, NH, United States

[2]Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States

Corresponding Author:
Nicholas C Jacobson, PhD
Center for Technology and Behavioral Health
Geisel School of Medicine
Dartmouth College
46 Centerra Parkway
Suite 300, Office # 333S
Lebanon, NH, 03766
United States
Phone: 1 6036467037
Email: Nicholas.C.Jacobson@dartmouth.edu

Abstract

Background: Social anxiety disorder is a highly prevalent and burdensome condition. Persons with social anxiety frequently avoid seeking physician support and rarely receive treatment. Social anxiety symptoms are frequently underreported and underrecognized, creating a barrier to the accurate assessment of these symptoms. Consequently, more research is needed to identify passive biomarkers of social anxiety symptom severity. Digital phenotyping, the use of passive sensor data to inform health care decisions, offers a possible method of addressing this assessment barrier.

Objective: This study aims to determine whether passive sensor data acquired from smartphone data can accurately predict social anxiety symptom severity using a publicly available dataset.

Methods: In this study, participants (n=59) completed self-report assessments of their social anxiety symptom severity, depressive symptom severity, positive affect, and negative affect. Next, participants installed an app, which passively collected data about their movement (accelerometers) and social contact (incoming and outgoing calls and texts) over 2 weeks. Afterward, these passive sensor data were used to form digital biomarkers, which were paired with machine learning models to predict participants' social anxiety symptom severity.

Results: The results suggested that these passive sensor data could be utilized to accurately predict participants' social anxiety symptom severity (r=0.702 between predicted and observed symptom severity) and demonstrated discriminant validity between depression, negative affect, and positive affect.

Conclusions: These results suggest that smartphone sensor data may be utilized to accurately detect social anxiety symptom severity and discriminate social anxiety symptom severity from depressive symptoms, negative affect, and positive affect.

(J Med Internet Res 2020;22(5):e16875) doi:10.2196/16875

KEYWORDS

biomarkers; machine learning; technology assessment, biomedical; social anxiety; social anxiety disorder; mobile phone

Introduction

Background

Social anxiety disorder (SAD) affects approximately 13% of Americans [1]. It is characterized by an intense, persistent, and exaggerated fear of evaluation or scrutiny in social situations and associated with behavioral avoidance [2]. SAD has a high socioeconomic cost, as it is associated with increased risk of school dropout, reduced productivity in the workplace, and lower quality of life [3]. Individuals with SAD symptoms are also at increased risk for developing depressive disorders [4], with comorbidity rates estimated between 30% and 70% in clinical and community samples [5-7].

Broadly speaking, anxiety around social situations (eg, fear of rejection or embarrassment) is relatively commonplace and can negatively impact individuals even outside the confines of this diagnostic category [8]. Furthermore, given the social inhibition and private anguish inherent to this pathology, SAD symptoms are often mistaken for shyness by others or perceived as a character flaw by the individual. As a result of these misconceptions, the nuanced effects of the condition are frequently underreported and underrecognized. Indeed, individuals with SAD often avoid consulting physicians about their psychological problems [9], and studies estimate that only 35% of individuals who meet clinical criteria receive treatment [3]. Thus, a sizable proportion of individuals who are struggling with SAD symptoms and could benefit from intervention go unaided. Moreover, there is a large time gap between disorder onset and the individual receiving treatment, with 36% experiencing SAD for more than 10 years before seeking help [10]. Currently, much of our understanding of the occurrence and presentation of SAD symptomatology is based on dispositional measures that have limited ecological validity. There is a great need for novel methodologies to improve our understanding and ability to identify individuals who may be vulnerable to developing this debilitating condition.

Smartphones have shown promise in recent years as ecologically valid tools for monitoring and predicting one's behavior and psychological state [11-14]. Specifically, passive usage and sensor data streams (eg, accelerometers, microphones, and GPS) offer insight into momentary behaviors that can serve as proxies for important mental health variables. The overarching goal of harnessing such data is to better understand, predict, and ultimately intervene when subtle behaviors are suggestive of problematic pathology. These data can be used to evaluate indicators of pathology over time with minimal burden or cost to the individual, researchers, or the broader health care system. Previous research examining patterns of smartphone use by socially anxious individuals have investigated a number of constructs including level of smartphone *addiction* [15,16], communication preferences (eg, preference for texting over voice calls) [17], behavioral markers before outgoing phone calls [18], location data (ie, using GPS) [19], and use of camera and health-related apps [20].

Previous Findings

Despite the promise of offering a better understanding of the contextual factors related to SAD symptom severity, most of the current research to date has not examined whether symptom severity can be accurately predicted utilizing only sensor data (ie, without additional severity indicators). This investigation is necessary to assess the utility of smartphone sensor data as a stand-alone predictive tool. Particularly, research is needed to determine whether passive sensor data could present a viable alternative to traditional assessments. Some research has examined the within-sample correlations of specific sensor metrics related to SAD symptom severity with significant correlations between some passive sensor data, including location, movement, calls, and texts [19,21]. However, correlations presented in previous publications have been based on general linear models utilizing the entire dataset and, consequently, may overfit the sample and overestimate how well these models would generalize to new independent samples [22]. In addition, the correlations used in these methods were not based on machine learning models, which can combine features to better predict SAD symptom severity (consequently, the absolute correlations between these features and social anxiety were only 0.01-0.36). The study by Boukhechba et al [21] was the only previous study that used the same publicly available data contained within this study.

To date, few researchers have investigated the out-of-sample accuracy of predicting social anxiety from these passive sensors [17,20]. Rather than examining SAD symptom severity continuously, one such study created three categories of SAD symptom severity (low, mid, and high) [18], which artificially and arbitrarily changes the scale and reduces the variance of the outcome [23]. Although previous researchers were successful in predicting social anxiety severity from phone calls, the out-of-sample prediction only explained 15.38% of the total variation in symptoms based on the statistics reported [19], indicating that the majority of the variance is explained by other variables not captured in the model. Consequently, although passive sensor data hold promise in assessing constructs related to SAD symptom severity, more research is needed to determine whether passive data can accurately predict symptoms to the degree that they represent a potential alternative to traditional assessments.

Study Aims

This study sought to test the utility of passive smartphone sensor data (ie, incoming and outgoing calls, text messages, and accelerometer data), gathered over 2 weeks via the Sensus mobile app as predictors of SAD symptoms in an unselected sample of undergraduates. Studies show that anxiety symptoms are common in undergraduate populations [24], and analog samples are useful for examining this pathology on a continuum [25]. Moreover, the study also utilized this sample to limit the influence of heterogeneity, given that undergraduates often experience similar environmental stressors and life phases. We hypothesized that we would be able to accurately predict SAD symptom severity with at least moderate accuracy (correlations greater or equal to 0.5 between out-of-sample predicted and observed social anxiety symptom severity) [26]. We also hypothesized that the predicted SAD symptom severity would show discriminant validity, evidenced by significantly higher correlations between the observed and predicted SAD symptom severity, compared with correlations with measures of affect (ie, depression symptoms, negative affect, and positive affect). Furthermore, given the novel nature of these data, we had some exploratory aims namely, we were interested in whether the data indicated any nuanced biomarkers that may be worth examining in future iterations of this work.

Methods

Participants

The current sample comes from a public use dataset [21]. A total of 72 undergraduate students within the United States consented to participate in the study (37/72, 51% female, mean age 19.8 years, SD 2.4; age range 18-23 years; 30/72, 41% white, 27/72, 37% Asian, 3/72, 4% black, 3/72, 4%

Latinx/Hispanic, and 9/72, 12% multiracial or unspecified). Participants were recruited through email advertisements sent to university email listserv for undergraduate students and the psychology department study participant pool. Participants were required to own their own Android devices (with an operating system 4.3 or higher). Although 72 participants were enrolled, only 59 with any phone calls, text data, and accelerometer data across the study period were a part of this study (all three channels were required to be included in the current analysis).

Measures

Social Interaction Anxiety Scale

The Social Interaction Anxiety Scale (SIAS) represented the primary outcome for this study [27]. The SIAS assesses the level of anxiety experienced in social situations using 20 self-report statements. Participants rate each of these items on a 0 (*not at all characteristic of me*) to 4 (*extremely characteristic of me*) Likert scale. In comparing patient populations with nonpatient populations, the SIAS demonstrated superior ability in differentiating SAD from nonanxious controls, with 82% to 86% sensitivity and 82% to 90% specificity using a cutoff point of 34 [28,29]. The SIAS also demonstrated high convergent validity (r=0.72) with the social phobia scale [28]. Previous research suggests that the optimal cutoff score for differentiating clinically significant SAD for the SIAS is 36 and that this cutoff point results in 93% sensitivity and 60% specificity in differentiating SAD from other clinically anxious groups, including panic disorder and agoraphobia [30]. Thus, the SIAS has strong convergent and discriminant validity in assessing a range of social anxiety symptoms. Moreover, the internal consistency of the SIAS in the current sample was good (alpha=.83) [11].

Depression, Anxiety, and Stress Scale—Depression Scale

The Depression, Anxiety, and Stress Scale (DASS-21) was used to assess depressive symptoms as a measure of discriminant validity for the study. The depression scale has excellent convergent validity (r=0.78 with the personal disturbance depression scale, r=0.66 with the Hamilton depression scale) [31] and adequate discriminant validity (r=0.62 for personal disturbance anxiety scale, r=0.59 for the Hamilton anxiety scale) [31]. The depression scale also has excellent internal consistency in the current sample (alpha=.91) [11].

Positive Affect Negative Affect Schedule

The positive affect negative affect schedule (PANAS) is a commonly utilized measure of positive and negative affect and was used in this study as a means of assessing discriminant validity [32]. The PANAS is a 20-item instrument that asks participants to rate the degree to which they experience positive (eg, *alert*, *inspired*, *enthusiastic*) and negative (eg, *distressed*, *upset*, *guilty*) affect *in general* on a 1 (*very slightly or not at all*) to 5 (*extremely*) Likert scale. The negative affect subscale of the PANAS has demonstrated convergent validity with the Beck Depression Inventory (r=0.58), the Hopkins Symptom Checklist (r=0.74), and the State Trait Anxiety Inventory state anxiety scale (A-State; r=0.51); while the positive affect subscale is negatively correlated with these measures (rs=−.35, −.19, and −.35, respectively [32]). Both subscales exhibit good internal consistency (alpha negative affect=.87, alpha positive affect=.88 [32]; note that the raw items were not reported for this sample, so the internal consistency of these scales is unknown for the current sample).

Passive Sensor Data

Accelerometer data were collected once per second (1/Hz) for the duration of the study. Notably, accelerometers sampled at this frequency have been used to detect related psychopathology (eg, major depression and bipolar disorders; pain and worry severity) [33-35]. Likewise, incoming and outgoing calls and text timestamps were recorded during the study period (note that the text message content was not included). Passive sensor data were collected for approximately 2 weeks (mean 16.41 days, SD 2.69).

Procedure

The study was approved by the institutional review board at the University of Virginia. Participants received partial course credit or payment for their participation, and they were told that the study examined how their thoughts and feelings interacted with their daily environment. They were also instructed on the type of data that would be collected from their mobile phones. Individuals provided informed consent and attended two laboratory visits. During the first laboratory visit, they completed the SIAS and other measures (not relevant to this study). After this, participants installed the Sensus mobile app onto their own Android phones. They returned approximately 2 weeks later, where they completed additional measures and were debriefed (although the authors may have collected these same measures during the follow-up assessment, only the baseline assessments were released).

Planned Analyses

Overview of Analyses

Before any analyses, a set of biomarkers was created for each person's data. Note that passive sensor data are based on within-person variation (ie, changes over time), and yet the primary hypothesis is based on a between-person question (ie, stable individual differences across people, which matches previous research conceptualizing anxiety disorders as dynamical systems) [36]. Consequently, the same set of biomarkers was extracted separately for each person from the time series of their passive sensor data, and, then, these biomarkers were used for interindividual analyses.

Accelerometer Biomarkers

The accelerometer data were processed consistently with a previously published procedure [33]. The first feature creation was used to describe the overall distribution of the outcome scores; this included the following raw data metrics: (1) the mean, (2) median, (3) mode, (4) minimum, (5) maximum, (6) skewness, (7) kurtosis, and (8) SD. Furthermore, we included (9) the root mean square of successive difference from 1 and 2 lags difference, and (10) the first through 99th quantiles in increments of 1 percentile. Thus, these features were predominantly created to extract relevant signals related to the functional form of the raw data.

The second set of features was created to represent the autoregressive dynamic nature of the raw data, which can measure complex relationships of temporal stability and oscillatory patterns while imposing a constraint for smoothness. This was constructed using the differential time-varying effect model, which is based on a generalized additive modeling framework utilizing the following formula for each person. In particular, the mean of $y_{i,l'}$, $\mu_i \equiv E(y_{i,l'})$ is linked to a semiparametric predictor, η_i, expressed as:

$$\eta_i = f_1(TD_{i,l})y_{i,l}$$

Here, y represents the raw data for each measurement i, at lag l. Note that l' represents a stacked vector of the outcome such that the same time series is stacked repeatedly to account for each potential lagged relationship l. The term f_1 represents a smooth based on a thin-plate regression spline. Note that the term f_1 is a nonparametric component wherein the effects of a series of covariates on the mean of the transformed dependent response variable are of unknown functional forms. TD reflects the time difference between the measurement occasion i and lag l. The primary term interest is thus: $f_1(TD_{i,l})y_{i,l}$, which is a varying-coefficient model representing the linear relationship between the lagged outcome of y on itself at later time points as a function of nonlinear time differences [37]. Features were created separately for each person's raw data (ie, one model per person). The extracted feature was the predicted varying-coefficients across the entire time series from all possible lags.

The third set of features was based on the spectral analysis, wherein each of the estimated power spectral densities was recovered for the raw data [38]. This represents the decomposed cyclical patterns that are common throughout the time series within the data. The power density describes the degree of strength of the variation in the raw data as a function of frequency. There was no missing accelerometer biomarker data.

Text Message Biomarkers

Notably, the text message data were much sparser than the accelerometer data, and consequently, fewer biomarkers could be derived. The distributional outcomes described within the accelerometer biomarkers above (eg, mean, median, mode, etc) were applied to the vector of the time difference between the text messages with the goal of processing the distribution of the time difference between adjacent text messages (attempting to capture whether the participant engaged in text messages with long delays or shorter-time periods). This was first applied for all text messages, then all incoming text messages, and, finally, all outgoing text messages. The typical distribution variance between persons was also captured by taking the same distributional features of the time differences for each person that the person contacted, and then calculating the SD of the distribution of time differences (note that this was to account for the potential that a socially anxious person may vary in the speed with which they text back persons with whom they are close as opposed to persons with whom they enjoy a more distant relationship). This resulted in a measure of how much variation there was in the length of time between contacts. Finally, the last feature that was extracted was the number of total texts. There was no missing text message biomarker data.

Call Biomarkers

The following features were created to process the call data: (1) the total number of calls, (2) the percentage of calls that were missed, (3) the percentage of calls that came during times the phone was idle (ie, to capture a time in which they might have been interrupted from their current tasks, which could facilitate avoidance among those with high social anxiety), (4) the number of total persons that the participants contacted, and (5) the distribution features extracted from the time differences between calls (ie., similar to the text messages, this approach was to account for how long it would take a person to call someone back, which we suspected could be higher in persons with high social anxiety). There was no missing call biomarker data.

Machine Learning

After each of the features were created, an ensemble of extreme gradient boosting machines (*XGboost*) was utilized. Extreme gradient boosting machines are learning algorithms composed of several weak tree-based learners. They have shown to be more robust at predicting outcomes compared with many traditional algorithms and are often the algorithm that consequently wins machine learning competitions for this reason. The lower level extreme gradient boosting learners were blocked in the following way: (1) predicting social anxiety from the accelerometer distribution features (which constituted one model), (2) predicting social anxiety from the accelerometer autoregressive dynamics (which comprised a second model), (3) predicting social anxiety from the accelerometer spectral densities (trained based on spectral densities in blocks of 1000 features each), (4) predicting social anxiety from the text message biomarkers, and (5) predicting social anxiety from the call biomarkers. Note that each of the extreme gradient–boosting models utilized leave-one-out cross-validation, such that the features that were extracted were the predictions from the model without including that participant in the model (ie, all features were out-of-sample predictions). Finally, the final ensemble model was trained based on the model predictions of the lower order models, with the higher-order model also being based on an extreme gradient boosting model. The final ensemble model also utilized leave-one-out cross-validation, such that the final model was based on an out-of-sample prediction.

Outcome Metrics

The primary outcome of interest was the correlation between the predicted and observed SAD symptom severity scores. In addition, the discriminant validity of the predicted SAD symptom scores was also compared by comparing the correlation between the SAD symptom scores based on the smartphone biomarkers and depression, negative affect, and positive affect. As there was missing data in the depression (5% missing), negative affect (14% missing), and positive affect (14% missing), multiple imputation was utilized to estimate the correlations for the discriminant validity. Similar to Ortiz et al [39], discriminant validity was compared by comparing the correlations between predicted SAD symptom severity based on the smartphone biomarkers and the observed SAD symptom

severity strength, compared with the correlation between the smartphone biomarkers and depression, negative affect, and positive affect, respectively [40].

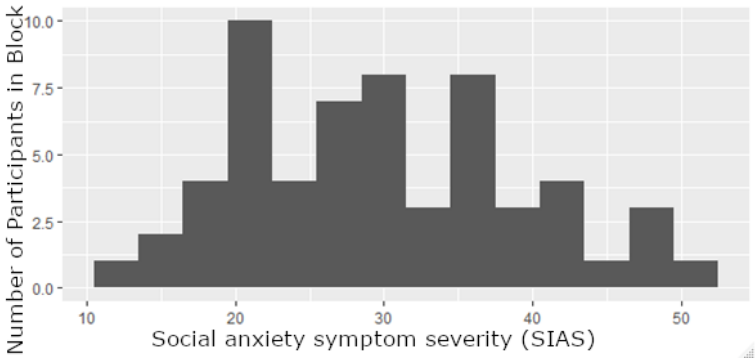
The variable importance from the model ensemble was identified, such that the most important feature will be extracted and plotted to determine the most important contribution to the model predictions. In addition, the t-distributed stochastic neighbor embedding (t-SNE) plot was examined from the lower order ensemble models and the final predictions to visualize the degree of separation between the predictors and social anxiety symptoms. The t-SNE plot is a visual depiction of how well the machine learning models naturally separate different degrees of SAD symptom severity. The variable importance of the primary feature was also extracted from the higher-order ensemble, as well as from the lower order model to determine the single most influential digital biomarker of the predictions of SAD symptom severity.

Results

Percentage of the Sample With Clinical Anxiety

The distribution of the SAD symptom severity is depicted in Figure 1. The mean of the SIAS was 29.125, with an SD of 9.407, and a range from 11 to 52. On the basis of previous validated cutoffs (score of 34 on the SIAS), 36% (21/59) of the sample was above the primary cutoff, suggesting they were at clinical levels of SAD. In applying the more conservative cutoff of 36 used in clinical outpatient samples to differentiate anxiety from other anxiety disorders, a total of 34% (21/59) of the sample was very likely to have clinically significant social anxiety (and not merely other anxiety disorders, based on the results of a previous study suggesting that this clinical cutoff discriminated SAD from other anxiety disorders [30]).

Figure 1. Social anxiety symptom severity based on the social interaction anxiety scale.



Primary Results

Convergent Validity

The results suggested that there was a strong correlation between predicted and observed social anxiety symptom severity r=0.702, 95% CI 0.543-0.812; P<.001 (see Figure 2 for a plot of the individual level predictions). Supporting our hypothesis, the bounds of the 95% CI suggested that the strength of the correlation was above 0.5. See Figure 3 for a visual depiction of the degree of separation of the lower order ensemble model features and SAD symptom severity.

Figure 2. Z-scores of the predicted social anxiety disorder (SAD) severity and the observed SAD symptom severity for each participant. Note that the patient index represents each of the 59 participants and not a continuous metric. SAD: social anxiety disorder; SIAS: social interaction anxiety scale.

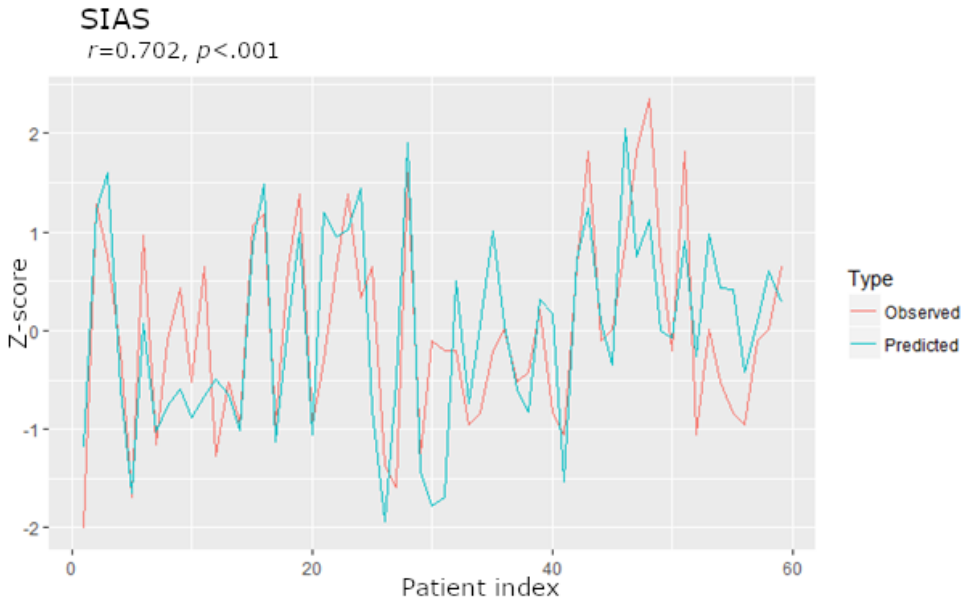
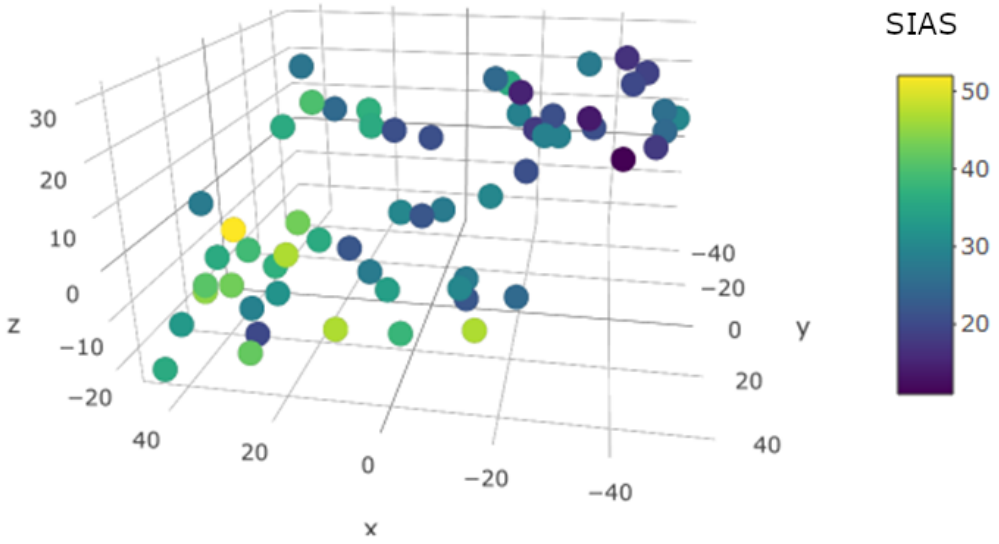


Figure 3. t-distributed stochastic neighbor embedding (t-SNE) plot depicting the ensembled model features (ie, the predictions extracted from the lower order ensembles) and the observed social anxiety symptoms. Note that closer points in the t-SNE are shown to be similar to one another, and dissimilar objects are shown to be farther away from one another. Note that the axes themselves are used to reduce the dimensionality of the machine learning features, so the Dimension 1, Dimension 2, and Dimension 3 labels are not of interest. This plot shows the ability to differentiate the level of social anxiety symptom severity based on the ensembled model features. As can be seen, the combination of the lower order ensembles was able to well differentiate SAD symptom severity. SAD: social anxiety disorder; SIAS: social interaction anxiety scale; t-SNE: t-distributed stochastic neighbor embedding.



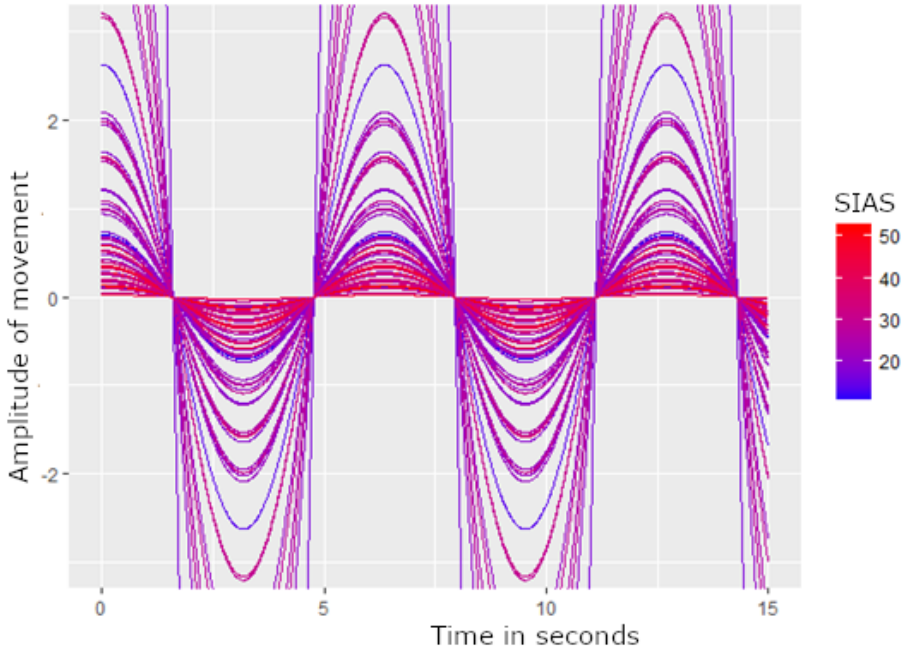
Discriminant Validity

Regarding the discriminant validity of the measure, the results suggested that there was a weaker but positive relationship between the predicted SAD symptom severity from smartphone biomarkers and depression severity (r=0.357, 95% CI 0.112-0.562; P=.005), and this correlation is significantly lower than that between the predicted SAD symptom severity from the smartphone biomarkers and the observed SAD symptom severity (Z=3.441; P<.001). Likewise, the predicted SAD symptom severity from smartphone biomarkers had a positive correlation with negative affect (r=0.384, 95% CI 0.143-0.583; P=.003), and the correlation is significantly lower than that between the predicted and observed SAD symptom severity (Z=3.484; P<.001). The predicted SAD symptoms had a nonsignificant negative correlation with positive affect (r=−0.138, 95% CI −0.380 to 0.122; P=.32), which was significantly lower than the correlation between predicted and observed SAD symptom severity (Z=5.980; P<.001). The magnitude of the variance explained was also more than three times greater between the predicted and observed SAD severity, compared with the variance explained by the predicted SAD severity and each of the discriminant constructs. This suggests that smartphone biomarkers demonstrate discriminant validity between depression, negative affect, and positive affect. Note that the partial correlation between the predicted SAD severity and the observed SAD severity was also still strong and significant when controlling for depression, positive affect, and negative affect (r=0.502, 95% CI 0.283 to 0.671; P<.001).

Exploratory Aims

Note that based on the variable importance metrics, the oscillations occurring every 6.35 seconds appeared to be the most influential contributor to the final ensemble model (see Figure 4). Graphical depictions of the results showed clear differences in amplitude between those with higher social anxiety symptoms and those with lower social anxiety symptom severity, such that those with low social anxiety tended to have greater oscillatory frequency patterns during this 6-second timespan.

Figure 4. Oscillations of movement at approximately every 6 seconds and the relationship to social anxiety symptom severity. Note that each line represents a separate person. The lines are shaded from blue (low social anxiety symptoms) to red (high social anxiety symptoms). This graph clearly depicts that those with high social anxiety symptoms tended to have lower amplitudes of these 6-second oscillations than those with higher social anxiety symptoms. SIAS: social interaction anxiety scale.



Discussion

Principal Findings

Our data indicated that calls, texts, and movement patterns captured by individuals' smartphones over 2 weeks provided sufficient information to predict severity of SAD symptoms with moderately strong accuracy (r=0.702), and that these patterns can also accurately discriminate social anxiety from depression (r=0.357), negative affect (r=0.384), or positive affect (r=−0.138). Although preliminary, these data are promising, as they suggest that simple behavioral information that is already passively collected for most individuals in the United States may represent a highly feasible, low-cost, low-burden, and specific mechanism for identifying people who are vulnerable to experiencing problematic levels of social anxiety.

Note that the present findings provide unique and incremental contributions of the previous public use dataset [21]. In particular, the previous research trained models based on the entire dataset and did not use a holdout sample, which may result in overfitting the sample and overestimate how well these models would generalize to new independent samples [22]. In addition, the correlations used via these methods were not based on machine learning models, which may combine features to better predict SAD symptom severity (consequently, the absolute correlations between these features and social anxiety were only 0.01-0.36) [21]. By utilizing a combination of features, rather than single features, the correlations were substantially better within the current sample despite being based on out-of-sample predictions.

The current findings are particularly notable when one considers the concurrent validity of SAD symptom measures. In particular, the convergent validity of these predictions is approximately equivalent to the convergent validity between established social anxiety symptom scales (eg, r=0.730 between the SIAS and the Liebowitz Social Anxiety Scale) [41]. This suggests that the accuracy of these behavioral phenotypes is approximately equivalent to the validity of symptom measures, which is particularly noteworthy given that self-report scales are known to tend to overly correlate with one another more strongly than other methods due to the artifact of shared method variance alone [42]. This has important implications for the field of psychology, as behavioral profiles are not based on subjective feelings, but rather replicable observable phenomena. Although this work is preliminary and more work is needed to develop a broader constellation of digital biomarkers, these behavioral profiles may be interesting outcomes to help organize and conceptualize psychiatric disorders themselves.

This study has many notable strengths. In particular, the smartphone app collected passive data from participants over 2 weeks continuously (once per second), allowing for rich behavioral signals that are well beyond the temporal precision available in existing social anxiety measures. Likewise, this study also applied some cutting-edge machine learning techniques in analyzing these digital biomarkers. This study also utilized leave-one-out cross-validation to examine the overall performance, which directly examines the degree to which the trained model generalizes to unseen data, and these models continued to suggest that they had high predictive accuracy. Finally, this study included a large percentage of those at clinical levels of SAD.

Limitations

This study has some limitations that provide direction for future research. First, our sample size was modest, and participants were all undergraduate students, which may limit the

generalizability of findings as SAD affects individuals of all ages and educational backgrounds. Furthermore, although 36% (20/59) of the sample reported symptom severity above the suggested clinical cutoff on the SIAS, we did not use a clinical sample in this study. Thus, it will be important that future replications of this work be conducted in larger, more representative samples. Larger samples would also allow for more power to test possible moderating variables (eg, gender). Second, passive data collection does not allow us to quantify the type of social contact that someone may be experiencing (ie, it may be difficult to infer whether someone is contacting a relative or a distant acquaintance). Third, the current models were trained to predict between-person differences, and, consequently, it is unknown whether the results would generalize to predicting within-person variability.

Our data also revealed that the most influential contributor to the predictive model was oscillatory frequency patterns during a 6.35-second timespan, such that less socially anxious individuals evidenced greater oscillatory frequency than high-social anxiety counterparts. Given the nature of the speed, we suspect that this reflects the length of several sequential strides during walking (ie, where this pace would reflect a slightly below average stride of six consecutive strides based on persons average walking pace) [43]. We suspect that this oscillatory pattern occurred as it best reflected continued walking frequency (ie, between 1 and 2 strides in a local environment likely has very high instability as persons only navigate very little distance). Nevertheless, given the novel nature of this metric, the true implications of this finding are unclear. It is possible that this pattern reflects persons with low social anxiety walking at a consistent confident and steady pace, whereas persons with high social anxiety walking might walk more quickly and less confidently or at an irregular pace. Continued research is needed to clarify the degree to which this finding is specific to this population, and the relevance of this metric for our understanding of symptom presentation.

Conclusions

Taken together, our study extends recent efforts to utilize passive smartphone sensor data to improve the field's ability to detect nuanced behavioral indicators of problematic pathology [44,45]. This is especially important for individuals with social anxiety, given that the occurrence of these symptoms is more frequent than is typically reported. However, this method of harnessing naturally occurring behavioral data is certainly relevant for identifying and better understanding a range of maladaptive thoughts and behaviors that underlie psychiatric conditions more broadly.

Conflicts of Interest

NJ is the owner of a free app entitled Mood Triggers. He does not receive any direct or indirect revenue from his ownership of the app (ie, the appl is free; there are no advertisements, and the data are only being used for research purposes). SW has received salary support from Telefonica Alpha, Inc. She is a presenter for the Massachusetts General Hospital Psychiatry Academy in educational programs supported through independent medical education grants from pharmaceutical companies. She has received royalties from Elsevier Publications, Guilford Publications, New Harbinger Publications, and Oxford University Press. She has also received speaking honoraria from various academic institutions and foundations, including the International Obsessive-Compulsive Disorder Foundation and the Tourette Association of America. In addition, she has received payment from the Association for Behavioral and Cognitive Therapies for her role as Associate Editor for the Behavior Therapy journal, as well as from John Wiley & Sons, Inc. for her role as Associate Editor for the journal Depression & Anxiety.

References

1. Kessler RC, Petukhova M, Sampson NA, Zaslavsky AM, Wittchen H. Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. Int J Methods Psychiatr Res 2012 Sep;21(3):169-184 [FREE Full text] [doi: 10.1002/mpr.1359] [Medline: 22865617]
2. American Psychiatric Association. Diagnostic And Statistical Manual Of Mental Disorders: DSM-5. Fifth Edition. Arlington, Virginia: American Psychiatric Publishing; 2013.
3. Ruscio AM, Brown TA, Chiu WT, Sareen J, Stein MB, Kessler RC. Social fears and social phobia in the USA: results from the National Comorbidity Survey Replication. Psychol Med 2008 Jan;38(1):15-28 [FREE Full text] [doi: 10.1017/S0033291707001699] [Medline: 17976249]
4. Jacobson NC, Newman MG. Anxiety and depression as bidirectional risk factors for one another: a meta-analysis of longitudinal studies. Psychol Bull 2017 Nov;143(11):1155-1200. [doi: 10.1037/bul0000111] [Medline: 28805400]
5. Adams GC, Balbuena L, Meng X, Asmundson GJ. When social anxiety and depression go together: a population study of comorbidity and associated consequences. J Affect Disord 2016 Dec;206:48-54. [doi: 10.1016/j.jad.2016.07.031] [Medline: 27466742]
6. Dalrymple KL, Zimmerman M. Treatment-seeking for social anxiety disorder in a general outpatient psychiatry setting. Psychiatry Res 2011 May 30;187(3):375-381. [doi: 10.1016/j.psychres.2011.01.004] [Medline: 21310497]
7. Kessler RC, Stang P, Wittchen HU, Stein M, Walters EE. Lifetime co-morbidities between social phobia and mood disorders in the US National Comorbidity Survey. Psychol Med 1999 May;29(3):555-567. [doi: 10.1017/s0033291799008375] [Medline: 10405077]

8. Furmark T, Tillfors M, Everz P, Marteinsdottir I, Gefvert O, Fredrikson M. Social phobia in the general population: prevalence and sociodemographic profile. Soc Psychiatry Psychiatr Epidemiol 1999 Aug;34(8):416-424. [doi: 10.1007/s001270050163] [Medline: 10501711]
9. Lecrubier Y, Wittchen HU, Faravelli C, Bobes J, Patel A, Knapp M. A European perspective on social anxiety disorder. Eur Psychiatry 2000 Feb;15(1):5-16. [doi: 10.1016/s0924-9338(00)00216-9] [Medline: 10713797]
10. Anxiety and Depression Association of America. Facts & Statistics URL: https://adaa.org/about-adaa/press-room/facts-statistics [accessed 2020-02-24]
11. Chow PI, Fua K, Huang Y, Bonelli W, Xiong H, Barnes LE, et al. Using mobile sensing to test clinical models of depression, social anxiety, state affect, and social isolation among college students. J Med Internet Res 2017 Mar 3;19(3):e62 [FREE Full text] [doi: 10.2196/jmir.6820] [Medline: 28258049]
12. Cornet VP, Holden RJ. Systematic review of smartphone-based passive sensing for health and wellbeing. J Biomed Inform 2018 Jan;77:120-132 [FREE Full text] [doi: 10.1016/j.jbi.2017.12.008] [Medline: 29248628]
13. Dogan E, Sander C, Wagner X, Hegerl U, Kohls E. Smartphone-based monitoring of objective and subjective data in affective disorders: where are we and where are we going? Systematic review. J Med Internet Res 2017 Jul 24;19(7):e262 [FREE Full text] [doi: 10.2196/jmir.7006] [Medline: 28739561]
14. Seppälä J, de Vita I, Jämsä T, Miettunen J, Isohanni M, Rubinstein K, M-RESIST Group, et al. Mobile phone and wearable sensor-based mhealth approaches for psychiatric disorders and symptoms: systematic review. JMIR Ment Health 2019 Feb 20;6(2):e9819 [FREE Full text] [doi: 10.2196/mental.9819] [Medline: 30785404]
15. Enez Darcin A, Kose S, Noyan CO, Nurmedov S, Yılmaz O, Dilbaz N. Smartphone addiction and its relationship with social anxiety and loneliness. Behav Inf Technol 2016 Apr 7;35(7):520-525. [doi: 10.1080/0144929X.2016.1158319]
16. Lee Y, Chang C, Cheng Z, Lin Y. How social anxiety and reduced self-efficacy induce smartphone addiction in materialistic people. Soc Sci Comput Rev 2016 Dec 29;36(1):36-56. [doi: 10.1177/0894439316685540]
17. Reid DJ, Reid FJ. Text or talk? Social anxiety, loneliness, and divergent preferences for cell phone use. Cyberpsychol Behav 2007 Jun;10(3):424-435. [doi: 10.1089/cpb.2006.9936] [Medline: 17594267]
18. Huang Y, Gong J, Rucker M, Chow P, Fua K, Gerber M. Discovery of Behavioral Markers of Social Anxiety from Smartphone Sensor Data. In: Proceedings of the 1st Workshop on Digital Biomarkers. 3089343: ACM; 2017 Presented at: DigitalBiomarkers'17; June 23, 2017; Niagara Falls, NY, USA p. 9-14. [doi: 10.1145/3089341.3089343]
19. Boukhechba M, Chow P, Fua K, Teachman BA, Barnes LE. Predicting social anxiety from global positioning system traces of college students: feasibility study. JMIR Ment Health 2018 Jul 4;5(3):e10101 [FREE Full text] [doi: 10.2196/10101] [Medline: 29973337]
20. Gao Y, Li A, Zhu T, Liu X, Liu X. How smartphone usage correlates with social anxiety and loneliness. PeerJ 2016;4:e2197 [FREE Full text] [doi: 10.7717/peerj.2197] [Medline: 27478700]
21. Boukhechba M, Daros A, Fua K, Chow P, Teachman B, Barnes L. DemonicSalmon: Monitoring mental health and social interactions of college students using smartphones. Smart Health 2018 Dec;9-10:192-203 [FREE Full text] [doi: 10.1016/j.smhl.2018.07.005]
22. Arlot S, Celisse A. A survey of cross-validation procedures for model selection. Statist Surv 2010;4:40-79. [doi: 10.1214/09-SS054]
23. Fitzsimons GJ. Death to dichotomizing. J Consum Res 2008 Jun 1;35(1):5-8. [doi: 10.1086/589561]
24. Blanco C, Okuda M, Wright C, Hasin DS, Grant BF, Liu S, et al. Mental health of college students and their non-college-attending peers: results from the national epidemiologic study on alcohol and related conditions. Arch Gen Psychiatry 2008 Dec;65(12):1429-1437 [FREE Full text] [doi: 10.1001/archpsyc.65.12.1429] [Medline: 19047530]
25. Stopa L, Clark DM. Social phobia: comments on the viability and validity of an analogue research strategy and British norms for the fear of negative evaluation questionnaire. Behav Cogn Psychother 2001 Oct 23;29(4):423-430. [doi: 10.1017/S1352465801004039]
26. Ferguson CJ. An effect size primer: a guide for clinicians and researchers. Prof Psychol Res Pract 2009 Oct;40(5):532-538. [doi: 10.1037/a0015808]
27. Mattick RP, Clarke JC. Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. Behav Res Ther 1998 May;36(4):455-470. [doi: 10.1016/s0005-7967(97)10031-6] [Medline: 9670605]
28. Brown EJ, Turovsky J, Heimberg RG, Juster HR, Brown TA, Barlow DH. Validation of the social interaction anxiety scale and the social phobia scale across the anxiety disorders. Psychol Assess 1997;9(1):21-27. [doi: 10.1037/1040-3590.9.1.21]
29. Heimberg RG, Mueller GP, Holt CS, Hope DA, Liebowitz MR. Assessment of anxiety in social interaction and being observed by others: The social interaction anxiety scale and the Social Phobia Scale. Behav Ther 1992;23(1):53-73. [doi: 10.1016/S0005-7894(05)80308-9]
30. Peters L. Discriminant validity of the social phobia and anxiety inventory (SPAI), the social phobia scale (SPS) and the social interaction anxiety scale (SIAS). Behav Res Ther 2000 Sep;38(9):943-950. [doi: 10.1016/s0005-7967(99)00131-x] [Medline: 10957828]
31. Crawford JR, Henry JD. The Depression Anxiety Stress Scales (DASS): normative data and latent structure in a large non-clinical sample. Br J Clin Psychol 2003 Jun;42(Pt 2):111-131. [doi: 10.1348/014466503321903544] [Medline: 12828802]

32. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: The PANAS scales. J Pers Soc Psychol 1988;54(6):1063-1070. [doi: 10.1037/0022-3514.54.6.1063]
33. Jacobson NC, Weingarden H, Wilhelm S. Digital biomarkers of mood disorders and symptom change. NPJ Digit Med 2019;2(1):3 [FREE Full text] [doi: 10.1038/s41746-019-0078-0] [Medline: 31304353]
34. Jacobson NC, Weingarden H, Wilhelm S. Using digital phenotyping to accurately detect depression severity. J Nerv Ment Dis 2019 Oct;207(10):893-896. [doi: 10.1097/NMD.0000000000001042] [Medline: 31596769]
35. Jacobson NC, O'Cleirigh C. Objective digital phenotypes of worry severity, pain severity and pain chronicity in persons living with HIV. Br J Psychiatry 2019 Jul 12:1-3 epub ahead of print. [doi: 10.1192/bjp.2019.168] [Medline: 31298167]
36. Frank B, Jacobson NC, Hurley L, McKay D. A theoretical and empirical modeling of anxiety integrated with RDoC and temporal dynamics. J Anxiety Disord 2017 Oct;51:39-46. [doi: 10.1016/j.janxdis.2017.09.002] [Medline: 28922648]
37. Jacobson NC, Chow S, Newman MG. The differential time-varying effect model (DTVEM): a tool for diagnosing and modeling time lags in intensive longitudinal data. Behav Res Methods 2019 Feb;51(1):295-315 [FREE Full text] [doi: 10.3758/s13428-018-1101-0] [Medline: 30120682]
38. Soica P, Moses R. Spectral Analysis of Signals. Upper Saddle River, New Jersey: Prentice Hall; 2004.
39. Ortiz A, Clinton A, Schaefer BA. Construct validity evidence for bracken school readiness assessment, third edition, Spanish form scores. Psychol Sch 2014 Dec 4;52(2):208-221. [doi: 10.1002/pits.21816]
40. Steiger JH. Tests for comparing elements of a correlation matrix. Psychol Bull 1980;87(2):245-251. [doi: 10.1037/0033-2909.87.2.245]
41. Heimberg RG, Horner KJ, Juster HR, Safren SA, Brown EJ, Schneier FR, et al. Psychometric properties of the Liebowitz social anxiety scale. Psychol Med 1999 Jan;29(1):199-212. [doi: 10.1017/s0033291798007879] [Medline: 10077308]
42. Orth U. How large are actor and partner effects of personality on relationship satisfaction? The importance of controlling for shared method variance. Pers Soc Psychol Bull 2013 Oct;39(10):1359-1372. [doi: 10.1177/0146167213492429] [Medline: 23798373]
43. Beauchet O, Annweiler C, Lecordroch Y, Allali G, Dubost V, Herrmann FR, et al. Walking speed-related changes in stride time variability: effects of decreased speed. J Neuroeng Rehabil 2009 Aug 5;6:32 [FREE Full text] [doi: 10.1186/1743-0003-6-32] [Medline: 19656364]
44. Wilhelm S, Weingarden H, Ladis I, Braddick V, Shin J, Jacobson N. Cognitive-behavioral therapy in the digital age: presidential address. Behav Ther 2020 Jan;51(1):1-14 [FREE Full text] [doi: 10.1016/j.beth.2019.08.001] [Medline: 32005328]
45. Burns MN, Begale M, Duffecy J, Gergle D, Karr CJ, Giangrande E, et al. Harnessing context sensing to develop a mobile intervention for depression. J Med Internet Res 2011 Aug 12;13(3):e55 [FREE Full text] [doi: 10.2196/jmir.1838] [Medline: 21840837]

Abbreviations

DASS-21: Depression, Anxiety, and Stress Scale
PANAS: positive affect negative affect schedule
SAD: social anxiety disorder
SIAS: social interaction anxiety scale
t-SNE: t-distributed stochastic neighbor embedding

Edited by G Eysenbach; submitted 04.11.19; peer-reviewed by M Hirschtritt, B Teachman; comments to author 04.01.20; revised version received 26.02.20; accepted 27.02.20; published 29.05.20.

Please cite as:
Jacobson NC, Summers B, Wilhelm S
Digital Biomarkers of Social Anxiety Severity: Digital Phenotyping Using Passive Smartphone Sensors
J Med Internet Res 2020;22(5):e16875
URL: http://www.jmir.org/2020/5/e16875/
doi:10.2196/16875
PMID:32348284

Original Paper

Exploring Strategies for Using Social Media to Self-Manage Health Care When Living With and Beyond Breast Cancer: In-Depth Qualitative Study

Cathy Ure[1], PhD; Anna Mary Cooper-Ryan[1], PhD; Jenna Condie[2], PhD; Adam Galpin[3], PhD

[1]Directorate of Allied and Public Health, School of Health and Society, University of Salford, Salford, Manchester, United Kingdom

[2]School of Social Sciences and Psychology, Western Sydney University, Sydney, Australia

[3]Directorate of Psychology and Sport, School of Health and Society, University of Salford, Salford, Manchester, United Kingdom

Corresponding Author:
Cathy Ure, PhD
Directorate of Allied and Public Health
School of Health and Society
University of Salford
Allerton Building
Salford, Manchester, M6 6PU
United Kingdom
Phone: 44 (0)161 295 5094
Email: c.m.ure1@salford.ac.uk

Abstract

Background: As breast cancer survival rates improve and structural health resources are increasingly being stretched, health providers require people living with and beyond breast cancer (LwBBC) to self-manage aspects of their care.

Objective: This study aimed to explore how women use and experience social media to self-manage their psychosocial needs and support self-management across the breast cancer continuum.

Methods: The experiences of 21 women (age range 27-64 years) were explored using an in-depth qualitative approach. The women varied in the duration of their experiences of LwBBC, which facilitated insights into how they evolve and change their self-management strategies over time. Semistructured interviews were analyzed inductively using a thematic analysis, a polytextual analysis, and voice-centered relational methods.

Results: The use of multiple social media platforms, such as YouTube, Facebook, WhatsApp, and Twitter, enabled women to self-manage aspects of their care by satisfying needs for timely, relevant, and appropriate support, by navigating identities disrupted by diagnosis and treatment and by allowing them to (re)gain a sense of control. Women described extending their everyday use of multiple platforms to self-manage their care. However, women experienced social media as both empowering and dislocating, as their engagement was impacted by their everyday experiences of LwBBC.

Conclusions: Health care professionals (HCPs) need to be more aware, and open to the possibilities, of women using multiple social media resources as self-management tools. It is important for HCPs to initiate value-free discussions and create the space necessary for women to share how social media resources support a tailored and timely self-managed approach to their unique psychosocial needs.

(J Med Internet Res 2020;22(5):e16902) doi:10.2196/16902

KEYWORDS

breast cancer; social media; internet; self-management; psychosocial health; survivorship

Introduction

Background

Breast cancer remains the most common type of cancer in women [1]. Owing to the improvements in early diagnosis, treatment, and an aging population [2,3], survivorship rates and life expectancy for women living with and beyond breast cancer (LwBBC) are increasing. However, as health care systems are increasingly stretched, with significant gaps developing in health care resource provision, increasing patients' abilities to engage

in positive self-management behaviors when living with long-term conditions [4,5] has become a global issue [6].

Self-management—defined as "awareness and active participation by the person in their recovery, recuperation, and rehabilitation, to minimize the consequences of treatment, promote survival, health and well-being" [7]—is reported to support patient empowerment, increase self-efficacy, and lead to behavioral changes, while reducing demands on health care resources [4,5,7-9]. However, Rogers et al [9] suggest that more attention needs to be placed on the "contexts, resources, practices, priorities, and networks of patients living with a chronic condition to identify the nuanced ways in which self-care support and long-term condition management can be integrated into the open systems of people's everyday lives."

Women LwBBC report many ongoing and unmet psychosocial needs [10,11], including pain, fatigue, fear of recurrence, lymphedema, and hair loss [12-19]. As these complex issues require ongoing support, attending to the everyday resources, practices, priorities, and patient networks that women engage in for their own self-care, could offer health care professionals (HCPs) insights for better health care outcomes.

Social media have the potential to support women LwBBC to manage aspects of their own self-care, including managing unmet psychosocial needs. Social media are defined as a group of Web-based apps that enable the creation and sharing of user-generated content [20]. They offer patients instant access to information and new connections through easy access to other users [21]. Social media have become taken for granted health information resources [22,23] and include, for instance, a cancer tag ontology of hashtags on Twitter that links patients, doctors, caregivers, and advocates. This includes a weekly tweet chat in relation to breast cancer—#bcsm (breast cancer social media) [24,25]. However, research on social media use by women LwBBC is limited to a small number of studies [24,26-39]. These studies have focused either on discrete platform use (eg, Facebook, Twitter) or have adopted methods such as secondary data analysis [26-38], which do not acknowledge the complex ways in which social media are used and experienced by women LwBBC. Furthermore, research on social media as a component of electronic health interventions [6] ignores the experiences of women using commercial social media platforms to create, use, maintain, and generate grassroots informational and storytelling spaces, such as blogs, and community spaces, such as Facebook groups, to support their own self-management. Facebook groups can be *open*, *closed*, or *secret*. *Open* means anyone can see the group, who is in it, and can join it. *Closed* groups can also be seen by anyone. The group's name, description, and member list are publicly visible; however, only those who have been invited to a closed group can see its posts. A *secret* group is not publicly visible. Women must ask to join secret breast cancer groups and are added by a moderator, and only group members can see group posts. As women are generating digitally mediated support together in Web spaces, understanding how women LwBBC use social media to support self-management and self-care practices can help target informational and psychosocial support more appropriately and provide useful information to HCPs about women's self-management practices.

Objective

This study aimed to explore how women use and experience social media to self-manage their psychosocial needs and support self-management across the breast cancer continuum.

Methods

Design

A qualitative study was used to explore women's use of social media to support self-management when LwBBC as we were unable to find previous studies that explored use of multiple social media for this purpose. Semistructured interviews, including visual methods were developed to gain rich, detailed data [40]. Photographs were introduced into the interview setting through photo elicitation (social support images provided by the research team, eg, friends being together, families, HCPs, and mobile technologies, eg, women using mobile phones or laptops) or photo production (photographs taken by women to communicate their experiences of LwBBC) to elicit or trigger conversation [41]. An interview guide was developed by the research team.

Sample and Sample Size

Women had to be 18 years or older, with a previous diagnosis of breast cancer. It was not a requirement that women had to use social media at the present time. We were interested in women's experiences of social media use at any time since diagnosis. We were mindful that some women may have used social media at some point but had not found it helpful. It was as important to capture these experiences. We were interested in how women's experiences of social media use varied in relation to time since diagnosis. We purposefully recruited women who had been LwBBC for less than twelve months, between 1 and 5 years, and 5 years or more, with equal numbers (n=7) recruited for each temporal period postdiagnosis. Participants emailed the lead researcher (CU) to register interest in participating after seeing study details online (in breast cancer Facebook groups [n=12], Twitter [n=2], charity websites [n=1]) or offline via cancer support centers (n=2), posters (n=1) or through word of mouth (n=3). In total, 44 women were interested in the study. Of the 44 women, 21 consented to participate, with 1 known to CU. Women decided which type of interview—photo elicitation or photo production—they wished to participate in.

Data Collection

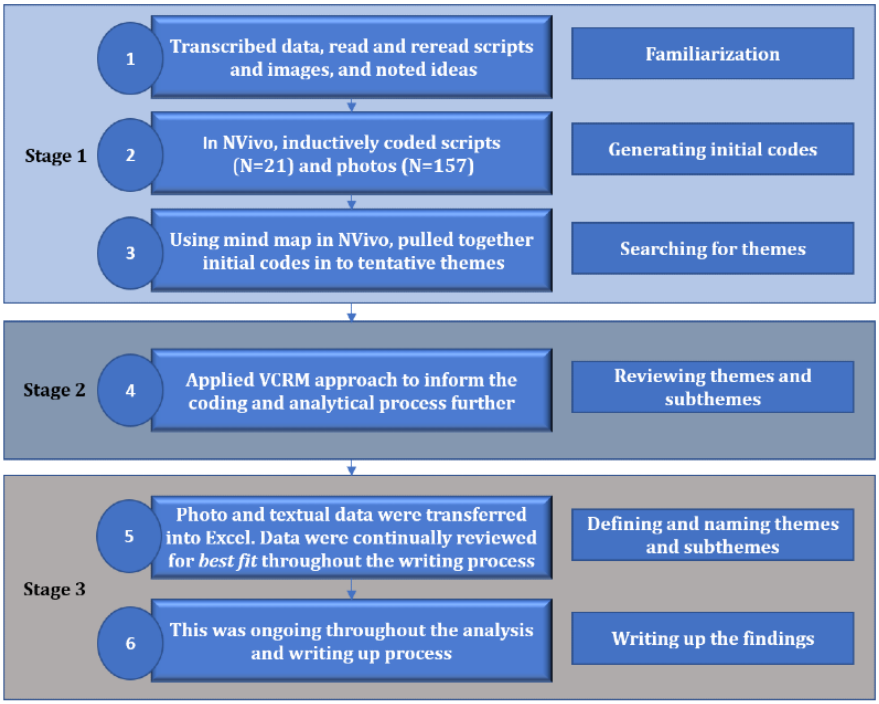
Face-to-face interviews carried out by CU were audio recorded and video recorded. The question structure was kept deliberately broad to enable women to have as much space as possible to explore their experiences of social media use. The first half of all interviews followed the same structure using the interview guide to ask broad questions related to personal experiences of breast cancer—women's overall social media use and use in relation to LwBBC. In the interviews using photo-production techniques, women then shared the photographs they had taken to discuss how they communicated their experiences of LwBBC with others. In the interviews, using photo elicitation techniques, Wortman's [42] study on social support was used to support probing questions. Photographs provided by the research team

were used to prompt responses related to possible providers of social support, including HCPs, work colleagues, friends and neighbors, family including children and parents, partners, peers, and service providers, for example, charities. Interviews took place in university or community settings. Field notes were written after each interview. Considerations about data saturation were guided by the concept of information power [43], that is, the more relevant information a sample holds, relevant to the study, the fewer participants are needed—researcher subjectivity [44] and taking a pragmatic approach to sample size [45], given the resources available. On the consent form, women were asked how they wanted quotes to be credited, that is, with their own name or a pseudonym. Most participants (16/21, 76%) waived their rights to anonymity. Pseudonyms are used for those who retained anonymity. Ethical approval was granted by the University of Salford (approval number: HSCR 15-71).

Data Analysis

Data were analyzed using thematic analysis [46], polytextual thematic analysis [47], and the voice-centered relational method [48] approaches. Figure 1 shows a review of the steps taken. Transcripts (including photographs) were inductively coded in NVivo (QSR International, version 11) by CU. Monthly review meetings were held with AMCR, AG, and JC to reflect on coding, to review field notes, and for thematic development.

Figure 1. Data analysis process flowchart. (VCRM: voice-centered relational method).



Participants' Review

Participants were invited to review the findings via email. Of 21 participants, 18 (86%) responded to the invitation and were forwarded the findings. Three photo-elicitation participants did not respond to the emailed invitations. The summary of the findings invited participants' responses by phone or email. Nine participants (9/18, 50%) responded. All (9) accepted the findings. Some offered comments about aspects of the findings that resonated with them. Others offered thanks for the opportunity to be involved.

Results

Characteristics of the Study Sample

A total of 21 women (age range: 27-64 years at the time of diagnosis) participated. Interviews lasted between 55 and 168 min (mean 99 min). The number of photographs (n=157) taken ranged from 3 to 47 (mean 17). Participants' characteristics are provided in Table 1.

Table 1. Participants' characteristics (N=21).

Characteristic	Values, n (%)
Age at time of diagnosis (years)	
<31	3 (14)
31-40	6 (29)
41-50	9 (43)
51-60	2 (9)
>61	1 (5)
Time since diagnosis	
<12 months	7 (33)
1-5 years	7 (33)
>5 years	7 (33)
Marital status	
Single	5 (24)
Cohabiting	2 (9)
Married	13 (62)
Divorced	1 (5)
Ethnicity	
White British	19 (90)
Black British	1 (5)
Mixed or multiple ethnicities	1 (5)
Employment status	
Full-time employment	7 (33)
Part-time employment	4 (19)
Unemployed	2 (9)
Retired	3 (14)
Student	2 (9)
Not working through choice	1 (5)
Unable to work due to health issues	5 (24)
Number of times diagnosed	
Once	17 (81)
Twice	4 (19)
Type of breast cancer	
Primary	11 (52)
DCIS[a]	4 (19)
Primary and DCIS	3 (14)
Local recurrence	2 (9)
Secondary	1 (5)
Treatments received	
Mastectomy	13 (62)
Lumpectomy	11 (52)
Chemotherapy	17 (81)
Radiotherapy	14 (66)
Tamoxifen	13 (62)

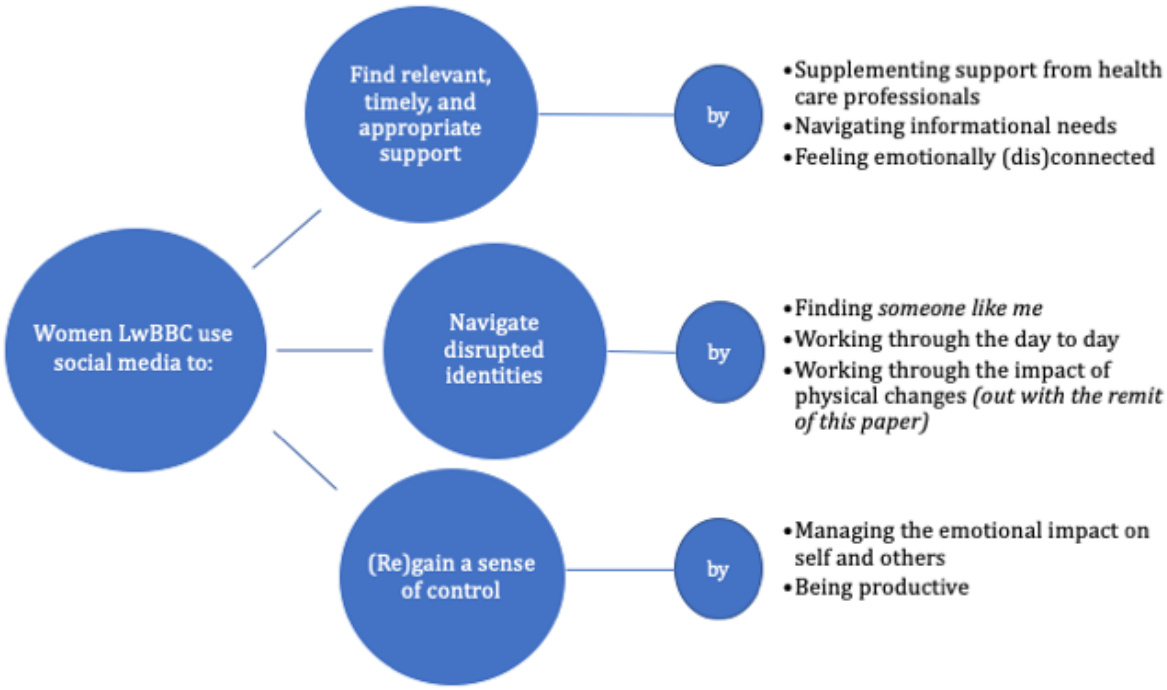
Characteristic	Values, n (%)
Type of interview	
Photo elicitation	12 (57)
Photo production	9 (43)
Most popular social media platform use in relation to breast cancer	
Facebook	17 (81)
YouTube	15 (71)
WhatsApp	8 (38)
Twitter	7 (33)

[a]DCIS: ductal carcinoma in situ.

Themes

A total of 3 main themes with 8 subthemes were identified in relation to why women use social media to support their experiences of LwBBC (Figure 2). For the purposes of this paper, the most relevant themes and subthemes that inform HCPs about women's use related to self-care and self-management were discussed. Working through the impact of physical changes, a subtheme identified as part of the wider PhD study, was, therefore, not discussed here.

Figure 2. Thematic map: women's use of social media to support self-management when living with and beyond breast cancer.



Theme 1: Finding Relevant, Timely, and Appropriate Support

Most women were active social media users at the time of their diagnosis. Extending their day-to-day use to find support was a logical extension of existing social media practices. Women described gaining support on social media through 3 subthemes: supplementing support from HCPs, navigating informational needs, and feeling emotionally (dis)connected.

Supplementing Support From Health Care Professionals

Women highly valued the clinical expertise provided by HCPs; however, collectively, they described themselves as cautious users of secondary health care provision. They avoided *mithering* or *bothering* breast care nurses, as *they never have enough time*. Phoning the breast care nurse was viewed as intrusive:

> *I don't want to be bothering them with phone calls when they are in clinic.* [Sarah, time since diagnosis <12 months]

Therefore, women who described using social media to gain support felt unwilling or unable to access their HCPs. Indeed, all women using social media described how easy access to experiential support from other women LwBBC reduced their sense of needing to access HCPs. Furthermore, sometimes women decided that they no longer needed to seek reassurance from their general practitioner:

> *I mean I tried to make appointments with my GP, but you know the way things are going with the NHS [National Health Service] and all that. It is like three weeks until my next appointment. I don't need to now. I just go on the group [Facebook group] and think' oh, OK, alright yeah.* [Jojo, time since diagnosis 1-5 years]

Instead, women can experience closed and secret groups on Facebook, as providing immediate, relevant responses that reassure and inform:

> *There would be people to talk to rather than having to ring up the nurse, leave a message on the breast cancer nurse line, feeling really bad because they are really busy, and they are running around doing other things and then waiting for them to ring me back perhaps that same day perhaps not that same day erm, to this was...this immediacy that, you know, we get used to with technology erm...* [Jayne, time since diagnosis 1-5 years]

Women at all stages of the breast cancer continuum described their social media use as supplementing professional support and as a way to gain agency by removing their reliance on HCPs as the source of all breast cancer knowledge. By using social media to address questions and concerns quickly, women voiced notions of feeling empowered.

Navigating Own Informational Needs

Initially, women used information searching as a coping strategy. For some, information searching began after diagnostic testing and before formal diagnosis. Women reported experiencing information overload in the clinical setting and used social media to fill knowledge gaps at key points in their *patient journey*, outside of the formal clinical encounter:

> *Because there's that much information, you can't possibly take it all in at the appointment. And when they tell you, you are almost kind of shocked anyway. So, you don't digest any of it. Nothing is retained. Absolutely nothing is retained.* [Nicola, time since diagnosis 1-5 years]

Women described *Googling* and seeking information from other women LwBBC on social media as common practice. Cancer charity websites, such as Cancer Research UK and Breast Cancer Care, were considered *legitimate*, *trustworthy*, and *up to date*. They were often women's first port of call. However, when women wanted further information relating to their own specific experiences, they described these sites as providing insufficient breadth or depth to satisfy their needs. In addition, some women experienced informational support from breast cancer charities as leaflet driven, which was misaligned with their everyday information-seeking practices. Women supplemented static Web- and leaflet-based information with active (eg, liking, commenting, sending messages) and passive consumption (eg, lurking) [49,50] of other women's experiences of LwBBC. Women described lurking in closed Facebook breast cancer groups immediately after diagnosis, which provided a *depth* of knowledge and built women's confidence to advocate for themselves:

> *You kind of come across a post and there'll be like 47 comments, by the time you've read all of that you've had quite an in-depth insight into that particular issue, so I just read a lot.* [Kirsty, time since diagnosis <12 months]

Using Twitter to follow and learn from other women, LwBBC was described as supporting joint decision making with HCPs:

> *I found Twitter really useful in that because then I started following lots of people, so by the time I spoke to people I was already pretty well informed, or I felt like I was anyway, um, obviously it gave me the opportunity to ask some questions then.* [Sheena, time since diagnosis <12 months]

In addition, women used YouTube to gain visual information relating to practical aspects of treatment and managing the effects of treatment, including lumpectomy, radiotherapy, and mastectomy procedures. Some women used YouTube to watch mastectomies *to see what they did* for greater knowledge and understanding.

Women supplemented Web-based information through active and passive consumption of experiential knowledge principally using Facebook, YouTube, and Twitter. By moving in and out of platforms and different groups on platforms, women gathered information at the appropriate time for them, determined by them. This supported women's ability to cope with the amount of information they encountered when newly diagnosed and supported adjustment and informed anticipation of what the next stage in their breast cancer experience entailed at different stages of the breast cancer continuum. By engaging in seeking, sifting, evaluating, and sharing information, women validated their experiences and were better equipped to advocate for themselves across a range of everyday settings.

Feeling Emotionally (Dis)connected

Some women used different social media platforms to navigate feelings of disconnection from other people to mediate relationships that provided emotional support and to connect with other women who shared similar experiences. Many described family and friends as *uncomfortable* because *they don't know what to say.* Some women talked about *my breast cancer* in everyday WhatsApp conversations. In this space, conversations about cancer were normalized, as women voiced their experiences while also providing family and friends with space and time to craft effective supportive responses. By publicly posting photographs as status updates on Facebook, women reported a sense that their story was being seen and heard. However, social media use was entangled in specific experiences. For instance, women showed highly tailored approaches to accessing emotional support at times, which were particularly challenging, such as the week of receiving chemotherapy. Women described these as important connections that reduced the sense of isolation and feelings of loneliness, which they felt unable to share with their family. Social media enabled connection with those who *understand* at a time when women felt disconnected from their normal support structures:

> *And those times when you are sat home for a week, bored out of your brains, feeling like death, it's quite*

> *nice to connect with somebody that's going through the same thing, yet you've not got the energy to talk so, you know what I mean, so it's been really good for that because I think I would have felt quite lonely…yeah.* [Sarah, time since diagnosis <12 months]

This emotional connectedness was described by 2 women as *a lifesaver*:

> *I cannot think of a single source that would provide even close to the amount of...even close to the amount of support the YBCN [Younger Breast Cancer Network] has provided for me. I never looked elsewhere.* [Delphi, time since diagnosis 1-5 years]

Many women preferred gaining emotional support through closed or secret Facebook groups or other platforms such as WhatsApp, Skype, and FaceTime, as these digital spaces supported intimate conversations, feelings of proximity, and "the reality of it [breast cancer] sometimes." Women described using different platforms simultaneously to scale how private or public they were about different aspects of their experiences. After finding social media groups or digital spaces that satisfied individual needs, some women developed personal relationships with other women LwBBC, which remained significant and important to them many years after their original diagnosis.

Theme 2: Navigating Disrupted Identities

Women described how breast cancer presented challenges to their sense of identity, which they navigated in numerous ways. The following 2 subthemes are relevant to HCPs: *finding someone like me* and *working through the day to day*.

Finding Someone Like Me

Postdiagnosis, many women used social media to find *someone like me*. Connecting with women in the *same boat*, who looked similar and were experiencing similar treatments helped reduce feelings of uncertainty. Women used different approaches to find women on social media, including looking for women of a similar age, with the same breast cancer type, and at the same stage *of the cancer journey.* Similarly, women targeted groups related to their experiences following particular treatments, including being *flat* postmastectomy, or having lymphedema. Most of the Facebook groups that women joined were grassroots, closed, and moderated:

> *A lot of us didn't want reconstruction and some were thinking about going flat completely and one of them mentioned the Flat Friends group cos I, I wear a (pause) prosthesis; I didn't have reconstruction I decided to join that group cos at one stage I thought erm do I go flat completely?* [Millie, time since diagnosis >5 years]

The need to find *similar others* on social media continued for some women along the breast cancer continuum. When women felt a difference between themselves and others, they sought out women in other groups they more closely identified with.

Some women identified difficulties with the notion of *someone like me* when they described experiencing the *hierarchy of suffering* [51], "whereby some kinds of suffering, pain, and misfortune are perceived as more difficult or signify a somewhat unique, superior source of suffering" (p.953). The *hierarchy of suffering* was described as invalidating personal experiences, and some women moved away from breast cancer discussions on Twitter and Facebook to actively manage their self-care. They described using social media to support their psychosocial health by extending friendships and interests with *similar others* with similar other interests, outside of breast cancer.

Working Through the Day to Day

Women at all stages of LwBBC described having to work through aspects of their breast cancer experience daily. For those with secondary breast cancer or further along the breast cancer continuum, many women detailed the impact and side effects of treatment as *constant reminders* and reported physical, emotional, psychological, and social challenges. Some women proactively used Twitter and Facebook to support themselves in the hospital setting while receiving treatment. This relieved pressure on the family to attend appointments, particularly for those living with secondary breast cancer:

> *I can't have somebody coming every three weeks with me, it's...who's got time? Who's got the energy? Who's got the effort? I don't mean that in a bad way, I know that it's a drag.* [Jo, time since diagnosis <5 years]

For Jo, working through the day to day involved using social media to extend the clinical encounter to her social networks. She used her iPad to connect with her Twitter followers during treatment, drawing on support in real time as and when she needed it.

For some women, *work* was experienced by managing side effects daily as a result of their treatment. Social media, including YouTube, were used to find solutions to alleviate the discomfort experienced, including managing lymphedema. However, regular emotional labor was also required to cope with anxiety created by social media posts when women discussed nonadherence to clinical guidelines. A total of 11 women, all at different stages of the breast cancer continuum, discussed issues with tamoxifen adherence. Women shared how nonadherence and being challenged about their own adherence created dissonance and anxiety for others:

> *And then I look on the [...] network and quite a few people say, "why are you having Tamoxifen? I'd put up a fight against that. I'm not on Tamoxifen I don't think it's a good idea. I'm having this drug instead and erm" so again that [….] is, now I am having a bit of a worry and a bit of a wobble about being on this Tamoxifen.* [Michelle, time since diagnosis <12 months]

Although some women felt conflicted when other women LwBBC challenged clinical guidelines, many women used social media as a tool to gain a sense of control.

Theme 3: (Re)gaining a Sense of Control

Women described (re)gaining a sense of control through 2 subthemes: managing the emotional impact on self and others and being productive**.**

Managing the Emotional Impact on Oneself and Others

Women described seeking to control the emotional impact of LwBBC on others by shielding them from aspects of LwBBC. Often, women made use of messaging services when initially diagnosed to inform others to (re)gain some control over disseminating their *news*. In the closed Facebook groups, women at all stages of LwBBC were able to have conversations, which they felt they could not have, or did not want to have, with close family and friends. They described how conversations with other women LwBBC enabled them to be *more honest*, able to process concerns without putting additional *burden* on loved ones, and which they had control over in terms of timing:

> *through social media I think you can be a little bit more honest because (pause) you've not got as much invested in their feelings. If you know, what I mean and they're going through it so you can't shock or scare them or make them feel (pause). There's no guilt in telling somebody on social media that yeah you do feel like shit, do you know what I mean, cos they're not going to come rushing round to your house, so there's that distance so I think you can definitely feel you can be more honest* [Sarah, time since diagnosis <12 months]

Women were purposeful in determining which platforms best supported their preferred communicative approaches and controlled when to publish personal information. WhatsApp provided women with a sense of intimacy, privacy, and connectedness, both with women LwBBC and their family/friends. Where women did not use WhatsApp and were not members of closed Facebook groups but were Facebook users, they posted to achieve responses that were (emotionally) manageable by controlling how they conveyed their experiences:

> *I would post "first out of six chemos. Last chemo – nailed it." That kind of thing. Erm, I do put it as very matter of fact. I did not say anything like "chemo is crap. I feel awful." And I would never post anything like erm, "I'm really down today" or anything like that. It was always very upbeat. I didn't want anyone to pity me.* [Wendy, time since diagnosis 1-5 years]

Women also reported having to learn to protect themselves emotionally when using social media, as sometimes content was experienced as threatening. Women reported anxiety—"you don't know what you're going to find"—when searching for content or reading about others' experiences. Women described strategies to control exposure to content so that it did not impinge negatively on their psychological health. This included prompt closing of content identified as having the *wrong atmosphere* and turning off push notifications from Facebook groups. Women described adopting flexible strategies of joining and leaving groups and conversations to reduce the negative impact on their sense of well-being across the breast cancer continuum. Any sense of information threat was met with a change in strategy, including avoidance and adaptation. Therefore, although women described social media as enabling them to compartmentalize and control aspects of their experiences, women also articulated the challenges of access to 24-hour information about breast cancer brings. Women emphasized the need to find ways to control access to other women's experiences, continual contact with *support* groups, and different types of content.

Being Productive

Women LwBBC reported using their experiences productively through their social media use. In one way some women (re)gained a sense of control by creating contemporary social media–based health resources. These resources were often borne out of the lack of service provision and included the development of new Web-based spaces including Twitter chats (#bbcww) and Facebook groups: for younger women with breast cancer, to support children of parents affected by cancer, and for women who wish to remain *flat.*

Some women over 12 months postdiagnosis felt *productive* by *giving back* through sharing personal experiences with women more recently diagnosed. However, many acknowledged this support as draining and adapted their level of involvement and exposure to content to meet their own self-care needs at any given point in time.

Similarly, engaging with oncologists on social media was seen to support women in making decisions about their own health care. Examples of successful advocacy were acknowledged by others LwBBC as something to *learn from:*

> *I've been able to then go to my Oncologist and say, "Look at this, this is what they're doing over in America, when are we getting it here?" Or "this is the treatment now available, when can I have it?"* [Jo, time since diagnosis >5 years]

For Jo, trying to influence other women LwBBC to develop positive attitudes toward improving their physical health provided purpose. By sharing updates on her own exercise goals, she sought to engage women in positive self-management behaviors. Jo used social media to inform, educate, and encourage others to be physically active to increase women's chances of accessing future treatments or surgical procedures through a focused approach to healthy, active living.

Women's photographs also demonstrated social media being used to actively challenge debilitating cancer narratives, which circulate in the mainstream press and on the Web. Countering problematic cancer narratives and having the right to reply was described as providing emotional release. Some women LwBBC, therefore, use social media as an opportunity for voice and reframing cancer conversations, which reduced their sense of disempowerment. Women can, therefore, develop complex social media identities that enable them to regain a sense of control through immersion in, shaping of, and sharing of expertise with others in ways that reciprocally supports their own individual needs.

Discussion

Principal Findings

This research identified 3 themes relating to women's use of social media to support self-management: *finding timely, relevant, and appropriate support* (support); *navigating*

disrupted identities (identity); and *(Re)gaining a sense of control* (control).

Women negotiate their entitlement to care [52] and use social media to supplement information provided by HCPs [53-55]. Significantly, women LwBBC at all stages of the breast cancer continuum are concerned about limiting demands on HCPs, for fear of *mithering* or *bothering*. This is supported by studies that show patients with cancer as reluctant to discuss psychosocial concerns with their clinicians [56] and users of online health communities (n=89) describing HCPs as *too busy for detailed discussions* [54]. To supplement the available support, some women are actively involved in digital labor in relation to meeting some of their psychosocial needs. Digital labor in a social media context relates to the unpaid creation of Web-based content and information [57]. In so doing, women articulate the perspective that they are reducing the potential demand for health resources. Rather than moving into new support environments postdiagnosis, such as using Web-based breast cancer charity forums, women's digital labor includes adapting and changing the *social* spaces they already use or occupy on social media, as demonstrated with the collaborative establishment of #bcsm [24,25] on Twitter. Evidence suggests that those who have diverse networks—characterized by numerous and varied network members (family, friends, acquaintances, and groups) who are in frequent contact with the individual—have better self-management capabilities among those with long-term conditions [58]. Furthermore, those with diverse networks use formal health care services less often than those in restricted, minimal family, family, or weak tie networks, presumably because of the increased number of connections [58]. Most often, the social spaces used are a logical extension of everyday use with women moving across platforms to suit their unique needs [59], in line with the trend for users to have accounts or profiles on multiple platforms [60-62]. Facebook, YouTube, WhatsApp, and Twitter were the most frequently used platforms in this study. Women adopt complex Web-based searching strategies to satisfy multiple needs, including timeliness (immediacy), relatability (women like me), and authenticity (experiential experts).

The first theme reports how women, by moving in and out of different groups and social media platforms, gain a sense of self-efficacy by gaining information at the appropriate time for them, determined by them. This supports women's ability to cope with the amount of information they encounter at challenging times along the breast cancer continuum and supports adjustment and informed anticipation of what the next stages in their cancer experiences entail. Using different social media platforms at different times demonstrates active and conscious decision making in tailoring connection and information seeking according to specific needs at any point in time.

Feeling an emotional connection to other women LwBBC was a significant factor in using social media, particularly when women felt disconnected from their usual support structures. Some women found that WhatsApp provided *mediated intimacy* [63] and supported the continuation of relationships postdiagnosis that felt disconnected or strained when face to face. By introducing everyday experiences of LwBBC into WhatsApp group conversations, women normalized conversations about cancer and succeeded in attending to their need to talk about their experiences. Women experienced WhatsApp as enabling message recipients to craft effective supportive responses.

The second theme captures the challenges for women in navigating disrupted identities and coping with *biographical disruption* [64-66]. Social media afford opportunities to link with and learn from *similar others*. When women experience a gap in service provision, they create or join niche groups [67] to find *likeminded* women. However, sometimes women find that participating in Facebook groups or on Twitter disrupts the validity of their experiences. They withdraw to better support their self-care. Women's experiences encapsulate the ongoing debate about the *internet's potential to create and diminish community* [68] with women experiencing participation as both positive and negative, sometimes simultaneously. More research is required to understand whether participation in social media conversations impacts adherence to clinical guidelines, including tamoxifen adherence and engagement in physical activities.

In the third theme—*(Re)gaining a sense of control*—women work to limit the *burden* of LwBBC on their families by self-managing aspects of emotional work through other women on social media. WhatsApp and Facebook Messenger were sometimes used following initial diagnosis to inform others. This was protective of psychosocial health by removing uncertainty around sharing the *diagnostic narrative* and afforded a sense of control through framing experience to influence the type of response required. On Twitter and in closed and secret groups, managing exposure to other women's experiences is challenging. Strategies employed to limit exposure to potentially threatening content support previous research [69] and align with women's approaches to information searching in relation to breast screening [70]. Similarly, push notifications on smartphones controlling the flow of information [71] from breast cancer groups acted as negative *constant reminders* of breast cancer. Being able to control information flow by changing notifications enabled a sense of control. The ability to use social media to compartmentalize the experiences of LwBBC was seen as a benefit of social media use.

Clinical Implications

Although evidence suggests that social media are shifting aspects of the patient-provider relationship [72], the original research aims of this study did not include generating outcomes for HCPs as a specific objective. However, the findings indicate that there are opportunities for HCPs and patients to work closer together to understand the benefits of social media use to support self-management. To encourage open conversations with patients, it may be useful for HCPs to encourage women LwBBC to share if and how they are addressing their self-management needs through their interactions with other women on social media. Given that many breast cancer groups on Facebook are closed or secret and WhatsApp groups are encrypted, there are challenges for researchers and HCPs alike in gaining awareness of what is shared in these spaces without dialog. Understanding which platforms or specific groups women find useful as self-management tools could enable

practitioners to signpost other women to social media resources that women find beneficial. Discussing the knowledge gained by women through platforms and social media groups provides opportunities for HCPs to support women's appraisal of information and could encourage open discussion about different approaches to self-management. Furthermore, there are opportunities for breast care nurses and oncologists to guest-moderate *chats* in grassroots social media spaces to support women's decision making and strategies for self-management.

Limitations

This was a qualitative study to understand the complexity of women's social media use when LwBBC. However, it did not provide insight into the relative extent of different aspects of social media use across a broad and representative population and did not provide insights into men's experiences of social media use when LwBBC. Further research should attempt to capture quantitative data to identify how social media use develops self-efficacy when LwBBC, supports self-management behaviors and impacts the overall sense of health and well-being across the cancer continuum.

Implications for Future Research

The ability to determine how, when, and where to access 24-hour support using social media provides opportunities for women globally to proactively engage in self-management practices unavailable a decade ago. Although women use social media in part to reduce demand on health care services, it is unknown whether use supports decision making or exacerbates issues within the clinical setting, for example, through poor decision making, which later increases demand for clinical services. Understanding HCPs' perceptions of the use of social media to support 24-hour self-managed care is an area for further research inquiry.

Acknowledgments

This study was funded as a doctoral study by the University of Salford. The authors would like to acknowledge Breast Cancer Care, Macmillan Cancer Support, and the grassroots Facebook groups—YBCN, Flat Friends, UK breast cancer group for survivors and sufferers, and the Word of Mouth Mammography e-Network for sharing details of this study to support recruitment and thank all the women who generously participated in this research.

Conflicts of Interest

None declared.

References

1. Bellanger M, Zeinomar N, Tehranifar P, Terry MB. Are global breast cancer incidence and mortality patterns related to country-specific economic development and prevention strategies? J Glob Oncol 2018 Dec;1(4):1-16. [doi: 10.1200/jgo.17.00207]
2. Jansana A, Posso M, Guerrero I, Prados-Torres A, del Cura MI, Castells X, et al. Health care services use among long-term breast cancer survivors: a systematic review. J Cancer Surviv 2019 Jun;13(3):477-493. [doi: 10.1007/s11764-019-00755-z] [Medline: 30968367]
3. Maddams J, Utley M, Møller H. Projections of cancer prevalence in the United Kingdom, 2010-2040. Br J Cancer 2012 Oct 25;107(7):1195-1202 [FREE Full text] [doi: 10.1038/bjc.2012.366] [Medline: 22892390]
4. Foster C, Fenlon D. Recovery and self-management support following primary cancer treatment. Br J Cancer 2011 Dec 8;105(Suppl 1):S21-S28 [FREE Full text] [doi: 10.1038/bjc.2011.419] [Medline: 22048029]
5. National Cancer Survivor Initiative (NCSI). 2010. National Cancer Survivorship Initiative Vision URL: http://www.ncsi.org.uk/wp-content/uploads/NCSI-Vision-Document.pdf [accessed 2018-02-07]
6. Triberti S, Savioni L, Sebri V, Pravettoni G. eHealth for improving quality of life in breast cancer patients: a systematic review. Cancer Treat Rev 2019 Mar;74:1-14. [doi: 10.1016/j.ctrv.2019.01.003] [Medline: 30658289]
7. Davies NJ, Batehup L. Towards a personalised approach to aftercare: a review of cancer follow-up in the UK. J Cancer Surviv 2011 Jul;5(2):142-151. [doi: 10.1007/s11764-010-0165-3] [Medline: 21253881]
8. Foster C, Calman L, Richardson A, Pimperton H, Nash R. Improving the lives of people living with and beyond cancer: generating the evidence needed to inform policy and practice. J Cancer Policy 2018 May;15:92-95 [FREE Full text] [doi: 10.1016/j.jcpo.2018.02.004]
9. Rogers A, Vassilev I, Sanders C, Kirk S, Chew-Graham C, Kennedy A, et al. Social networks, work and network-based resources for the management of long-term conditions: a framework and study protocol for developing self-care support. Implement Sci 2011 May 29;6:56 [FREE Full text] [doi: 10.1186/1748-5908-6-56] [Medline: 21619695]
10. Burg MA, Adorno G, Lopez ED, Loerzel V, Stein K, Wallace C, et al. Current unmet needs of cancer survivors: analysis of open-ended responses to the American Cancer Society Study of Cancer Survivors II. Cancer 2015 Mar 15;121(4):623-630 [FREE Full text] [doi: 10.1002/cncr.28951] [Medline: 25581252]
11. Park BW, Hwang SY. Unmet needs of breast cancer patients relative to survival duration. Yonsei Med J 2012 Jan;53(1):118-125 [FREE Full text] [doi: 10.3349/ymj.2012.53.1.118] [Medline: 22187241]

12. Aaronson NK, Mattioli V, Minton O, Weis J, Johansen C, Dalton SO, et al. Beyond treatment-psychosocial and behavioural issues in cancer survivorship research and practice. EJC Suppl 2014 Jul;12(1):54-64 [FREE Full text] [doi: 10.1016/j.ejcsup.2014.03.005] [Medline: 26217166]
13. Butt Z, Rosenbloom SK, Abernethy AP, Beaumont JL, Paul D, Hampton D, et al. Fatigue is the most important symptom for advanced cancer patients who have had chemotherapy. J Natl Compr Canc Netw 2008 May;6(5):448-455 [FREE Full text] [doi: 10.6004/jnccn.2008.0036] [Medline: 18492460]
14. Cheng KK, Wong WH, Koh C. Unmet needs mediate the relationship between symptoms and quality of life in breast cancer survivors. Support Care Cancer 2016 May;24(5):2025-2033. [doi: 10.1007/s00520-015-2994-0] [Medline: 26530229]
15. Dua P, Heiland MF, Kracen AC, Deshields TL. Cancer-related hair loss: a selective review of the alopecia research literature. Psychooncology 2017 Apr;26(4):438-443. [doi: 10.1002/pon.4039] [Medline: 26594010]
16. Fiszer C, Dolbeault S, Sultan S, Brédart A. Prevalence, intensity, and predictors of the supportive care needs of women diagnosed with breast cancer: a systematic review. Psychooncology 2014 Apr;23(4):361-374. [doi: 10.1002/pon.3432] [Medline: 24677334]
17. Foster C, Wright D, Hill H, Hopkinson J, Roffe L. Psychosocial implications of living 5 years or more following a cancer diagnosis: a systematic review of the research evidence. Eur J Cancer Care (Engl) 2009 May;18(3):223-247. [doi: 10.1111/j.1365-2354.2008.01001.x] [Medline: 19432917]
18. Glare PA, Davies PS, Finlay E, Gulati A, Lemanne D, Moryl N, et al. Pain in cancer survivors. J Clin Oncol 2014 Jul 1;32(16):1739-1747 [FREE Full text] [doi: 10.1200/JCO.2013.52.4629] [Medline: 24799477]
19. Trusson D, Pilnick A. The role of hair loss in cancer identity: perceptions of chemotherapy-induced alopecia among women treated for early-stage breast cancer or ductal carcinoma in situ. Cancer Nurs 2017;40(2):E9-16. [doi: 10.1097/NCC.0000000000000373] [Medline: 27070222]
20. Kaplan AM, Haenlein M. Users of the world, unite! The challenges and opportunities of social media. Bus Horiz 2010 Jan;53(1):59-68. [doi: 10.1016/j.bushor.2009.09.003]
21. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 May 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
22. Office for National Statistics. 2018. Internet Access: Households and Individuals, Great Britain: 2017 URL: https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2017#household-internet-access-continues-to-rise [accessed 2018-02-07]
23. Prescott J, Mackie L. 'You sort of go down a rabbit hole...you're just going to keep on searching': a qualitative study of searching online for pregnancy-related information during pregnancy. J Med Internet Res 2017 Jun 5;19(6):e194 [FREE Full text] [doi: 10.2196/jmir.6302] [Medline: 28583906]
24. Attai DJ, Cowher MS, Al-Hamadani M, Schoger JM, Staley AC, Landercasper J. Twitter social media is an effective tool for breast cancer patient education and support: patient-reported outcomes by survey. J Med Internet Res 2015 Jul 30;17(7):e188 [FREE Full text] [doi: 10.2196/jmir.4721] [Medline: 26228234]
25. Katz MS, Anderson PF, Thompson MA, Salmi L, Freeman-Daily J, Utengen A, et al. Organizing online health content: developing hashtag collections for healthier internet-based people and communities. JCO Clin Cancer Inform 2019 Jun;3:1-10 [FREE Full text] [doi: 10.1200/CCI.18.00124] [Medline: 31251658]
26. Bender JL, Katz J, Ferris LE, Jadad AR. What is the role of online support from the perspective of facilitators of face-to-face support groups? A multi-method study of the use of breast cancer online communities. Patient Educ Couns 2013 Dec;93(3):472-479. [doi: 10.1016/j.pec.2013.07.009] [Medline: 23928354]
27. Coll-Planas G, Visa M. The wounded blogger: analysis of narratives by women with breast cancer. Sociol Health Illn 2016 Jul;38(6):884-898. [doi: 10.1111/1467-9566.12405] [Medline: 26893129]
28. Crannell WC, Clark E, Jones C, James TA, Moore J. A pattern-matched Twitter analysis of US cancer-patient sentiments. J Surg Res 2016 Dec;206(2):536-542. [doi: 10.1016/j.jss.2016.06.050] [Medline: 27523257]
29. de Boer M, Slatman J. Blogging and breast cancer: narrating one's life, body and self on the internet. Womens Stud Int Forum 2014 May;44:17-25. [doi: 10.1016/j.wsif.2014.02.014]
30. de la Torre-Díez I, Díaz-Pernas FJ, Antón-Rodríguez M. A content analysis of chronic diseases social groups on Facebook and Twitter. Telemed J E Health 2012;18(6):404-408. [doi: 10.1089/tmj.2011.0227] [Medline: 22650380]
31. Doh H, Pompper D. Beyond the wounds: the SCAR Project as space for examining breast cancer and mastectomy experiences. Soc Semiotics 2015 Jun 5;25(5):597-613. [doi: 10.1080/10350330.2015.1041792]
32. Foley N, O'Mahony M, Lehane E, Cil T, Corrigan M. A qualitative content analysis of breast cancer narratives hosted through the medium of social media. Br J Med Med Res 2015 Jan 10;6(5):474-483. [doi: 10.9734/bjmmr/2015/15309]
33. Freedman RA, Viswanath K, Vaz-Luis I, Keating NL. Learning from social media: utilizing advanced data extraction techniques to understand barriers to breast cancer treatment. Breast Cancer Res Treat 2016 Jul;158(2):395-405 [FREE Full text] [doi: 10.1007/s10549-016-3872-2] [Medline: 27339067]
34. Himelboim I, Han JY. Cancer talk on Twitter: community structure and information sources in breast and prostate cancer social networks. J Health Commun 2014;19(2):210-225. [doi: 10.1080/10810730.2013.811321] [Medline: 24111482]

35. Kim E, Hou J, Han JY, Himelboim I. Predicting retweeting behavior on breast cancer social networks: network and content characteristics. J Health Commun 2016;21(4):479-486. [doi: 10.1080/10810730.2015.1103326] [Medline: 27007166]
36. Tan ML, Kok K, Ganesh V, Thomas SS. Patient information on breast reconstruction in the era of the world wide web. A snapshot analysis of information available on YouTube.com. Breast 2014 Mar;23(1):33-37. [doi: 10.1016/j.breast.2013.10.003] [Medline: 24215982]
37. Nzali MD, Bringay S, Lavergne C, Mollevi C, Opitz T. What patients can tell us: topic analysis for social media on breast cancer. JMIR Med Inform 2017 Jul 31;5(3):e23 [FREE Full text] [doi: 10.2196/medinform.7779] [Medline: 28760725]
38. Tsuya A, Sugawara Y, Tanaka A, Narimatsu H. Do cancer patients tweet? Examining the Twitter use of cancer patients in Japan. J Med Internet Res 2014 May 27;16(5):e137 [FREE Full text] [doi: 10.2196/jmir.3298] [Medline: 24867458]
39. Harris LN, Cleary EH, Stanton AL. Project connect online: user and visitor experiences of an internet-based intervention for women with breast cancer. Psychooncology 2015 Oct;24(9):1145-1151 [FREE Full text] [doi: 10.1002/pon.3734] [Medline: 25521661]
40. Silver J. Visual methods. In: Willig C, editor. Introducing Qualitative Research In Psychology. Third Edition. Maidenhead, UK: Open University Press; 2013:156-165.
41. Prosser J. Visual methodology: towards a more seeing approach. In: Denzin N, Lincoln Y, editors. The SAGE Handbook of Qualitative Research. Thousand Oaks, CA: Sage Publications; 2011:479-497.
42. Wortman CB. Social support and the cancer patient. Conceptual and methodologic issues. Cancer 1984 May 15;53(10 Suppl):2339-2362. [doi: 10.1002/cncr.1984.53.s10.2339] [Medline: 6367944]
43. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res 2016 Dec;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]
44. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qual Res Sport Exerc Health 2019 Jun 13;11(4):589-597. [doi: 10.1080/2159676x.2019.1628806]
45. Braun V, Clarke V. To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. Qual Res Sport Exerc Health 2019 Dec 26:- epub ahead of print. [doi: 10.1080/2159676x.2019.1704846]
46. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
47. Gleeson K. Polytextual thematic analysis for visual data. In: Reavy P, editor. Visual Methods in Psychology: Using and Interpreting Images in Qualitative Research. Abingdon, United Kingdom: Routledge; 2011:314-329.
48. Brown LM, Gilligan C. Meeting at the Crossroads. Garden Street, Cambridge, MA: Harvard University Press; 1992.
49. Trifiro BM, Gerson J. Social media usage patterns: research note regarding the lack of universal validated measures for active and passive use. Soc Media Soc 2019 May 25;5(2):205630511984874. [doi: 10.1177/2056305119848743]
50. Verduyn P, Ybarra O, Résibois M, Jonides J, Kross E. Do social network sites enhance or undermine subjective well-being? A critical review. Soc Issues Policy Rev 2017 Jan 13;11(1):274-302. [doi: 10.1111/sipr.12033]
51. Sidenius A, Mogensen O, Rudnicki M, Møller LM, Hansen H. Feeling lucky: hierarchies of suffering and stories of endometrial cancer in a Danish context. Sociol Health Illn 2019 Jun;41(5):950-964. [doi: 10.1111/1467-9566.12875] [Medline: 30740754]
52. Sulik GA. On the receiving end: women, caring, and breast cancer. Qual Sociol 2007 Mar 6;30(3):297-314. [doi: 10.1007/s11133-007-9057-x]
53. Ho Y, O'Connor BH, Mulvaney SA. Features of online health communities for adolescents with type 1 diabetes. West J Nurs Res 2014 Oct;36(9):1183-1198 [FREE Full text] [doi: 10.1177/0193945913520414] [Medline: 24473058]
54. Rupert DJ, Moultrie RR, Read JG, Amoozegar JB, Bornkessel AS, O'Donoghue AC, et al. Perceived healthcare provider reactions to patient and caregiver use of online health communities. Patient Educ Couns 2014 Sep;96(3):320-326. [doi: 10.1016/j.pec.2014.05.015] [Medline: 24923652]
55. Smailhodzic E, Hooijsma W, Boonstra A, Langley DJ. Social media use in healthcare: a systematic review of effects on patients and on their relationship with healthcare professionals. BMC Health Serv Res 2016 Aug 26;16:442 [FREE Full text] [doi: 10.1186/s12913-016-1691-0] [Medline: 27562728]
56. Ozakinci G, Swash B, Humphris G, Rogers SN, Hulbert-Williams NJ. Fear of cancer recurrence in oral and oropharyngeal cancer patients: an investigation of the clinical encounter. Eur J Cancer Care (Engl) 2018 Jan;27(1):-. [doi: 10.1111/ecc.12785] [Medline: 29024186]
57. Fuchs C. Social Media: A Critical Introduction. London, UK: Sage Publications; 2014.
58. Vassilev I, Rogers A, Kennedy A, Wensing M, Koetsenruijter J, Orlando R, et al. Social network type and long-term condition management support: a cross-sectional study in six European countries. PLoS One 2016;11(8):e0161027 [FREE Full text] [doi: 10.1371/journal.pone.0161027] [Medline: 27536988]
59. Harkin L. University of Central Lancashire. 2016. Exploring Visitors Experiences of Online Cancer Communities URL: http://clok.uclan.ac.uk/16646/1/16646%20Lydia%20Harkin%20%20Final%20e-Thesis%20%28Master%20Copy%29%20June%202016.pdf [accessed 2018-02-12]
60. Greenwood S, Perrin A, Duggan M. Pew Research Center. 2016 Nov 11. Social Media Update 2016 URL: http://www.pewinternet.org/2016/11/11/social-media-update-2016/ [accessed 2018-03-01]

61. Ofcom. 2016 Apr. Adults' Media Use and Attitudes: Report 2016 URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0026/80828/2016-adults-media-use-and-attitudes.pdf [accessed 2016-07-14]
62. Ofcom. 2017 Jun. Adults' Media Use and Attitudes: Report 2017 URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0020/102755/adults-media-use-attitudes-2017.pdf [accessed 2020-03-26]
63. Vetere F, Gibbs M, Kjeldskov J, Howard S, Mueller FF, Padell S, et al. Mediating Intimacy: Designing Technologies to Support Strong-Tie Relationships. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2005 Presented at: CHI'05; April 2 - 7, 2005; Oregon, Portland, USA p. 471-480 URL: http://dl.acm.org/citation.cfm?id=1055038 [doi: 10.1145/1054972.1055038]
64. Hubbard G, Forbat L. Cancer as biographical disruption: constructions of living with cancer. Support Care Cancer 2012 Oct;20(9):2033-2040. [doi: 10.1007/s00520-011-1311-9] [Medline: 22076620]
65. Liamputtong P, Suwankhong D. Breast cancer diagnosis: biographical disruption, emotional experiences and strategic management in Thai women with breast cancer. Sociol Health Illn 2015 Oct;37(7):1086-1101. [doi: 10.1111/1467-9566.12282] [Medline: 25922881]
66. Trusson D, Pilnick A, Roy S. A new normal?: women's experiences of biographical disruption and liminality following treatment for early stage breast cancer. Soc Sci Med 2016 Mar;151:121-129. [doi: 10.1016/j.socscimed.2016.01.011] [Medline: 26802368]
67. Bender JL, Jimenez-Marroquin MC, Ferris LE, Katz J, Jadad AR. Online communities for breast cancer survivors: a review and analysis of their characteristics and levels of use. Support Care Cancer 2013 May;21(5):1253-1263. [doi: 10.1007/s00520-012-1655-9] [Medline: 23179491]
68. Batenburg A, Das E. Virtual support communities and psychological well-being: the role of optimistic and pessimistic social comparison strategies. J Comput-Mediat Comm 2015 Aug 24;20(6):585-600. [doi: 10.1111/jcc4.12131]
69. Corter AL, Speller B, Sequeira S, Campbell C, Facey M, Baxter NN. What young women with breast cancer get versus what they want in online information and social media supports. J Adolesc Young Adult Oncol 2019 Jun;8(3):320-328. [doi: 10.1089/jayao.2018.0125] [Medline: 30648917]
70. Galpin A, Meredith J, Ure C, Robinson L. 'Thanks for letting us all share your mammogram experience virtually': developing a web-based hub for breast cancer screening. JMIR Cancer 2017 Oct 27;3(2):e17 [FREE Full text] [doi: 10.2196/cancer.8150] [Medline: 29079555]
71. Kim S, Kim S, Kang H. An analysis of the effects of smartphone push notifications on task performance with regard to smartphone overuse using ERP. Comput Intell Neurosci 2016;2016:5718580 [FREE Full text] [doi: 10.1155/2016/5718580] [Medline: 27366147]
72. Huo J, Desai R, Hong Y, Turner K, Mainous AG, Bian J. Use of social media in health communication: findings from the health information national trends survey 2013, 2014, and 2017. Cancer Control 2019;26(1):1073274819841442 [FREE Full text] [doi: 10.1177/1073274819841442] [Medline: 30995864]

Abbreviations

bcsm: breast cancer social media
HCP: health care professional
LwBBC: living with and beyond breast cancer

Edited by G Eysenbach; submitted 05.11.19; peer-reviewed by L O'Riordan, R Vue, K McCausland; comments to author 12.12.19; revised version received 05.02.20; accepted 21.02.20; published 25.05.20.

Please cite as:
Ure C, Cooper-Ryan AM, Condie J, Galpin A
Exploring Strategies for Using Social Media to Self-Manage Health Care When Living With and Beyond Breast Cancer: In-Depth Qualitative Study
J Med Internet Res 2020;22(5):e16902
URL: http://www.jmir.org/2020/5/e16902/
doi:10.2196/16902
PMID:32364510

Original Paper

Exploring the Characteristics and Behaviors of Nurses Who Have Attained Microcelebrity Status on Instagram: Content Analysis

Hanna Kerr[1], RN, MSCN; Richard Booth[1], RN, PhD; Kimberley Jackson[1], RN, PhD

Western University, London, ON, Canada

Corresponding Author:
Hanna Kerr, RN, MSCN
Western University
Room 3306, FIMS & Nursing Building, Western University
1151 Richmond St
London, ON, N6A 5B9
Canada
Phone: 1 2262390224
Email: hkerr4@uwo.ca

Abstract

Background: Instagram is a social media platform that enables users to share images and videos worldwide. Some nurses have used Instagram to document their experiences as a nurse and have subsequently gained microcelebrity status—that is, a user who purposefully seeks to amass a substantive Web-based following and has become recognized as a niche area of interest.

Objective: This study aimed to identify the characteristics and behaviors of microcelebrity nurses who act as influencers on Instagram and use their nursing profile to gain attention and presence on the Web.

Methods: A qualitative, exploratory, nonparticipatory content analysis of media and text generated by a purposeful sample of 10 registered nurses who use Instagram and sustain a definable microcelebrity status was conducted. In this study, manifest and latent data were examined to gain an understanding of the characteristics and behaviors of nurses who have attained microcelebrity status on Instagram.

Results: Data analysis revealed 5 themes of Instagram posts: (1) engaging Instagram users, (2) educational opportunities and insights, (3) nursing-related humor, (4) emotions experienced by nurses, and (5) media and narratives including patient details or work context. Messages were primarily positive in nature; however, multiple potential privacy, ethical, and professional issues were noted throughout the posted content.

Conclusions: The findings of this study help to expand the current knowledge related to the use of social media platforms such as Instagram, especially in regard to the emergence of nurses who use this form of technology to achieve or maintain a microcelebrity status. This study calls for additional research on nurses' attainment of microcelebrity status on social media as well as further policy development to adequately prepare nurses to navigate social media.

(J Med Internet Res 2020;22(5):e16540) doi:10.2196/16540

KEYWORDS

nursing; social media; professionalism; microcelebrity; Instagram; policy; influencer

Introduction

Significance

Every month, more than one billion people use the photo- and video-sharing social media platform, Instagram [1]. As a mobile photo-sharing platform, Instagram users can post photos and videos that are shared with other Instagram users, customized through the use of image processes (ie, filters), textual captioning up to 2200 characters in length, and hashtags [2]. Instagram users interact with others through a dynamic process of *liking* and *commenting* on the posted media as well as *following* fellow users. The number of followers an Instagram user has is publicly displayed on a user' s account page, regardless of the user's privacy settings [3].

Owing to the interactive nature of Instagram and the ability to identify the number of followers a specific user possesses, a behavior demonstrated by some Instagram users of purposefully cultivating and amassing a substantial Web-based following has become a recognized phenomenon, that is, if a social media user gains both substantial attention and a distinct following

within their group of followers, they are said to have attained *influencer* or *microcelebrity* status on social media [4-8]. Many individuals, including some nurses, have begun to purposefully leverage the connective powers of Instagram to generate microcelebrity persona through their Web-based presence.

At present, there are numerous Instagram accounts that exclusively feature content related to the nursing profession that has amassed sizable followings of more than 10,000 users, arguably pushing them toward microcelebrity status on Instagram [5,6,8]. However, given the potentially sensitive nature of the nursing role and related patient care, the use of Instagram by some users to post content related to the nursing profession is an area in need of deeper exploration. For instance, sharing of elements related to client care details and personal health information with the public raises potential concerns about the professional, ethical, and legal ramifications of such accounts and their related users.

Previous research examining the motivations of Instagram users and their use of social media platforms has found aspects related to (1) attention seeking; (2) social support; (3) belonging; (4) social interaction; (5) documentation; (6) learning about other users' lives; and (7) self-expression, as being important to users [9-13]. Within the nursing literature, several reasons have been explored regarding why nurses use social media platforms such as Instagram. The use of social media by nurses may assist in providing benefits for both nurses and patients, in part because of the ability to rapidly share information, assist users who are geographically isolated, and engage in health care planning [14-24].

Although social media have several potential benefits for nurses, challenges have also been identified with social media use in the profession. Social media posts can be disseminated to vast audiences immediately, which can have unintended results, especially if posts lack sensitivity or breach personal health information [16,17,20,22,23,25-27]. Given the subtle and encompassing nature of social media platforms to provide users with a false sense of security and privacy, the risk for health care providers to unintentionally cross professional boundaries on the Web is a contemporary reality for all nurses [17,18,23,25,28,29].

The consequences for nurses who breach patient privacy or confidentiality on social media are significant, including possible job termination, investigation and reprimand from nursing board(s), monetary fines, and loss of licensure. Furthermore, nurses may face civil or criminal charges, resulting in a potential jail time [17-20,28]. In recent years, there have been several instances of nurses violating the Health Insurance Portability and Accountability Act (HIPAA) policies on social media and being disciplined for their actions [20,26,27]. In Texas, a pediatric nurse posted comments on Facebook about a patient with measles who she cared for. No advertent personal information was disclosed; however, the nurse was terminated for HIPAA violation [26].

As a result of ongoing concerns related to unprofessional social media use, nurses are advised to use social media with caution and awareness regarding what they are posting [17,19,20,28]. Nurses are prohibited from posting any images or videos depicting patients or any content that would enable a patient to be recognized. In addition, nurses are warned that it is possible to violate privacy policies without explicitly disclosing patient information [17,19,20,28]. In fact, most instances of nurses violating patient privacy and confidentiality on social media are accidental in nature [17,28]. Finally, nurses are advised to refrain from interaction with patients over social media or use substantial caution in doing so [17,18,28,30-32].

Although the current body of literature examining social media usage by nurses contains insights into both positives and negatives of social media usage, there is an absence of research exploring the growing presence of microcelebrities in the profession who use platforms such as Instagram to showcase their interpretation of the profession. To proactively ensure the safety and privacy of patients and health care professionals in the coming years, it is necessary to examine the current characteristics and behaviors of nurse influencers who use Instagram to expose various elements of the nursing profession, especially those with significant Web-based followings.

Purpose of the Study

The purpose of this study was to identify the characteristics and behaviors of nurses who have attained microcelebrity status on Instagram and use their nursing profile to gain attention and presence on the Web.

Methods

Overview

This study was an exploratory, nonparticipatory content analysis of publicly available data found on Instagram, guided by Graneheim and Lundman's qualitative content analysis framework [33]. The framework developed by Graneheim and Lundman was selected for the study to enable researchers to gain a rich understanding of information by analyzing both manifest and latent data [33]. In this study, data were examined to identify the characteristics and behaviors of *microcelebrity Instagram users* [33]. The underlying meaning of posts was examined to gain an understanding of the types of messages being projected by microcelebrity Instagram users.

An initial step in Graneheim and Lundman's content analysis framework is the identification of the unit of analysis [33]. In this study, the unit of analysis was a textual description of the microcelebrity Instagram users' Instagram posts, including the direct caption and the researcher's description of each post. Once created, the unit of analysis was divided to form meaning units. The meaning units for this study were written descriptions of each individual Instagram post and the exact accompanying caption. The meaning units were condensed to summarize the manifest content of each Instagram post. After the manifest content emerged, abstraction took place, in which the summarized meaning units were further analyzed for their latent content and labeled with a code. Codes were grouped into subcategories and categories [33]. The categories were reviewed and revised, as emergent themes were identified [33].

Ethics Approval

The Western University Health Sciences Research Ethics Board deemed this study did not require ethical clearance for completion.

Sample

Purposive sampling was used to enable variation in identifying the characteristics and behaviors of nurses with active microcelebrity status on Instagram. To be included in the study, the following needed to be met: (1) the user's Instagram account had to be publicly available, possessing a minimum of 10,000 followers; (2) the 10 most recent posts from the user must have been created between 2017 and 2018; (3) all content should be written in English; (4) the majority of the user's content had to be focused on the nursing profession; and (5) the user must have self-identified as a registered nurse in their Instagram account description. Owing to the exploratory nature of this study, the sample was limited to registered nurses. A small sample enabled researchers to gain a rich understanding of the phenomenon within this group.

A priori, it was estimated that analysis of 10 to 15 microcelebrity Instagram accounts would result in data saturation. The amount of data to be collected from 10 to 15 microcelebrities (with 10 posts per individual) was deemed to be sufficient and comparable with other studies that used purposeful samples of social media data [34-36].

To obtain the sample, hashtag and keyword searches were performed on Instagram using the following search terms: *#nurse(s)*, *#registerednurse*, *#rn* as well as *nurse(s)*, *registerednurse*, and *rn*. These terms were selected to aid the identification of relevant Instagram accounts. When a hashtag is searched on Instagram, results consist of publicly available Instagram posts that have been tagged with the hashtag. Posts are displayed as *top posts* and *most recent* posts. *Top posts* populate hashtags that are trending, displaying 9 of the most liked posts containing that hashtag [3]. For each search, the 9 *top posts* for every hashtag were selected, and the Instagram user who created the post was identified and examined for study eligibility. Sampling was limited to the *top posts* search results, as the study aimed to identify Instagram users with active microcelebrity status.

Following the hashtag search process, an Instagram keyword search was conducted to identify additional eligible Instagram users. When an Instagram keyword search is conducted, the results populate, as a list of Instagram accounts that are relevant to the keyword searched. The first 10 results of each keyword search were assessed for study eligibility. Through the sampling process, 10 Instagram accounts were selected for the study. To protect the privacy of Instagram users, each account was randomly assigned a participant code number between 01 and 10.

Data Collection

The data collection and analysis framework for the study was developed from Graneheim and Lundman's [33] framework for qualitative content analysis. The data for this study consisted of each of the 10 microcelebrity Instagram user's 10 most recent posts (images, videos, and captions). To collect data, the researcher examined each Instagram user's posts individually. All study data were collected between September 2017 and August 2018. Each microcelebrity Instagram user was identified, and their 10 most recent posts were labeled in reverse chronological order. For each Instagram post, the researcher copied and pasted the entire caption to a secured word document. Then, the researcher described the manifest content of the post's photo or video in the written text. For example, the manifest content of microcelebrity Instagram user 10's, post 7 was captured as follows:

> *A collage of three images of the microcelebrity Instagram user. Image one: Microcelebrity Instagram user walking towards a retail store with a shopping cart. Image two: Microcelebrity Instagram user sleeping in a bed. Image three: Microcelebrity Instagram user sitting on a couch holding a glass of wine, an ice cream carton, and a pizza box. The heading of the collage states, "What nurses do on their days off..." The caption states, "This is my life! ... Please tell me I'm not alone."* [Microcelebrity Instagram user 10, Post 7]

The data collection process was completed for the 10 most recent posts of each of the 10 Instagram users, totaling 100 Instagram posts for analysis.

Data Analysis

From the unit of analysis, meaning units were created by dividing the text into each microcelebrity Instagram user's content and further subdividing them into individual Instagram posts. Thus, the meaning units for this study comprised a written description of each Instagram post. The meaning units were then condensed to summarize the manifest content of each Instagram post [33]. Through this process, the researcher abridged the direct meaning of both the photo or video and caption into a shortened textual description.

Following summarization, meaning units were exported to NVivo 12 (QSR International) for further analysis. Then, the process of abstraction took place, in which the summarized meaning units were analyzed for their latent meaning and labeled with a code. Once each post was coded, the codes were analyzed for similarities and differences and grouped into subcategories and then into categories. The categories were then reviewed, discussed, and revised. Once the revision process was complete, emergent themes were identified [33].

Results

Overview

Through data analysis, 5 themes emerged: (1) *engaging Instagram users*, (2) *educational opportunities and insights*, (3) *nursing-related humor*, (4) *emotions experienced by nurses*, and (5) *media and narratives including patient details or work context*. Seven of the microcelebrity Instagram users appeared to be female, and 3 microcelebrity Instagram users appeared to be male. Three of the microcelebrity Instagram users stated that they live in the United States of America, whereas the

geographic locations of the remaining microcelebrity Instagram users were unknown.

Engaging Instagram Users

Commonly, microcelebrity Instagram users interacted with followers by generating discussion within their posts. Multiple microcelebrity Instagram users posted questions related to the nursing profession or clinical knowledge and encouraged responses in the post's comment field. For example, microcelebrity Instagram users encouraged discussion related to dealing with a negative workplace culture and strategies to find a new nursing job. In addition, microcelebrity Instagram user 05 created multiple problem-solving posts that encouraged users to test nursing-related knowledge (Microcelebrity Instagram user 05, posts 2 and 10).

Along with a generic discussion of the profession or clinical knowledge, most microcelebrity Instagram users also created personalized brands based on some elements of their unique physical, cultural, or personality characteristics (eg, gender, ethnicity, and religion) and reinforced elements of this personalized brand within their Instagram posts. For example, microcelebrity Instagram user 04 repeatedly engaged in self-branding by highlighting the personal characteristics that made him unique. Microcelebrity Instagram user04 stated:

> *When people say, "What's a nurse?" They think of Florence Nightingale, a female wearing a little hat, white dress with a stethoscope. They don't automatically think of someone who's a Bilingual/Hispanic Male or a different culture ethnicity all Tatted covered with Ink...thats a MAN of GOD.* [Post 7]

Multiple microcelebrity Instagram users engaged in other branding activities by encouraging users to purchase individualized products that they developed or endorsed. Furthermore, promoting external merchandise and products was common, as 8 of the microcelebrity Instagram users promoted at least one merchandise or product that was not their own creation. For instance, microcelebrity Instagram users promoted businesses selling nursing-related uniforms, equipment, or services as well as non-nursing-related companies (eg, Cherokee Uniforms, Peloton, and Johnson & Johnson). In all these examples, it was unclear if the microcelebrity Instagram users were compensated for creating posts.

Educational Opportunities and Insights

Nursing-related educational opportunities and insights were a prominent theme of Instagram posts. Microcelebrity Instagram users encouraged followers to attend nursing-related educational events and posted photos of themselves at nursing conferences. In addition to promoting formal educational opportunities, multiple microcelebrity Instagram users shared their insights related to clinical nursing skills. For example, microcelebrity Instagram user 05 posted medication dosage questions, along with information about the medication, and encouraged users to complete the problem and comment on the solution. In addition, microcelebrity Instagram users provided their opinions on the various roles that nurses undertake. For example, microcelebrity Instagram user 04 posted 3 video clips in which operating room nurses discuss their responsibilities. In the videos, operating rooms were shown to appear to be in use, including patients on operating tables. It is unclear if the videos are real or performed; however, the videos appear realistic. Finally, only 2 posts related to obtaining formal nursing education at an undergraduate or master's level, both of which shared a general motivational message about persevering through the challenges associated with nursing school.

Nursing-Related Humor

In this study, the use of humor was evident in several Instagram posts. Frequently, humor was used in a manner that may only be understood by those with the contextual knowledge that comes with working as a health care professional, such as:

> *Remember don't over resuscitate your self with turkey and beer or you might end up with Abdominal Compartment Syndrome!* [Microcelebrity Instagram user 03, Post 3]

Another example was related to the experiences that nurses may encounter at the workplace. Microcelebrity Instagram user 01 posted an image of a white background with overlaying text:

> *"What were his respirations?" First of all, I haven't counted respirations since nursing school.* [Post 8]

The caption reads:

> *Too slow, too fast, normal, prepare to intubate. Those are the rates.* [Microcelebrity Instagram user 01, Post 8]

To fully understand these forms of humor, readers must possess the requisite knowledge of medical terminology as well as the contextual knowledge associated with working as a health care provider.

Humor was observed in multiple Instagram posts, as microcelebrity Instagram users described potentially stressful situations. Microcelebrity Instagram user 08 posted a series of memes, depicting a professional basketball player holding his arms out, mouth open, appearing shocked, and disgruntled during a game. The headings were as follows:

> *When ICU refuses to take report because the patient is "too unstable."* [Microcelebrity Instagram user 08, Post 5]
>
> *When your A&Ox4 patient decides they're too tired so they just pee in the stretcher.* [Microcelebrity Instagram user 08, Post 5]

The caption further read:

> *I wish I could say all of the above scenarios are made up-but no no, they're all real life.* [Microcelebrity Instagram user 08, Post 5]

Similar to the previous examples, users require contextual and clinically based knowledge to interpret the meaning of this post.

Emotions Experienced by Nurses

Inspiration, motivation, and encouragement were the most prominent emotions displayed by Instagram users, as every microcelebrity Instagram user shared at least one post demonstrating these emotions. Several posts were positive in

nature, as microcelebrity Instagram users presented optimistic captions or portrayed nurses as heroic figures. For example, microcelebrity Instagram user 02 expressed gratitude by stating:

> *There are many forms of family...For me, my work family is as much a part of my life as my real one...There's something about the few quiet moments on a morning like this before the chaos starts when you're reminded to give thanks for countless reasons.* [Post 9]

In addition to the positive emotions associated with the nursing profession, microcelebrity Instagram users described the challenges they experienced while nursing. The challenges described by microcelebrity Instagram users included mental and physical exhaustion, learning how to prioritize tasks, manage time, deal with workplace bullying, and shiftwork. Although many challenges associated with nursing were identified, every Instagram post that discussed a challenge concluded with a positive message. For example, microcelebrity Instagram user 02 stated:

> *Today was, legitimately, a "Top 3 Worst Shifts I've Ever Had as an ICU Nurse" kind of day. Nothing went right...& things just got worse by the hour...I'm exhausted...But no matter how bad the day gets—& trust me, in our world, it can get ugly—only your work family knows how to make you crack a smile...I genuinely believe that the long, ugly days & crazy hours impact how you see the world. Smile because you have love. Smile because you have life. Smile because life is precious, & every moment with loved ones matters...Give thanks, even after a long & torrid day.* [Post 8]

In these posts, microcelebrity Instagram users consistently shared messages of encouragement, despite experiencing challenging situations associated with the nursing profession.

Media and Narratives Including Patient Details or Work Context

Multiple Instagram posts depicted scenarios in which the photo or video appeared to have taken place in a workplace setting or potentially include patient information. For instance, multiple Instagram posts appear to depict patients in the hospital, although the sources of the images are not cited. microcelebrity Instagram user 03 posted an image of a patient's abdomen undergoing a surgical procedure, providing education on abdominal compartment syndrome in the caption. It is unclear if the patient consented to have this photo taken or if consent was obtained to post the image on Instagram. In another post, microcelebrity Instagram user 03 posted an image of a patient sitting in a stretcher, covered in dried blood, smiling, and surrounded by hospital staff members. The caption was unrelated to the image, asking volunteers for a trauma course. In addition, microcelebrity Instagram user 04 posted a series of video clips showing what appeared to be multiple patients on operating tables while receiving preprocedure treatments. No qualification was made in the post as to whether the patients depicted in the image were actually real or standardized patient actors.

Furthermore, several of the microcelebrity Instagram users' Instagram posts consisted of images that appear to have been taken in workplaces, with no patients seen. Multiple posts depicted microcelebrity Instagram users dressed in work attire while in hospital settings; other posts showed patient care areas with no patients present.

Discussion

Principal Findings

The findings of this study have identified a gap between nursing research, policy, and practice. Previous research has recommended the integration of social media with the nursing profession; however, policies continue to advise nurses to keep their professional lives separate from any social media use [14,28,37-40]. Further inquiry should be conducted to determine why this gap exists and identify how elements of policy and practice can be integrated in a safe and efficient manner.

All 10 microcelebrity Instagram users attempted to engage with followers, which supports previous research suggesting that a primary motive for Instagram use is social interaction [9]. Eight microcelebrity Instagram users directly participated in self-branding or merchandise solicitation. In particular, microcelebrity Instagram user 04 consistently promoted their personal brand, attempting to generate a distinct persona through reinforcement of this brand to their followers. It appears that microcelebrity Instagram user 04 may have been attempting to use his charismatic nature to help foster a sizable Web-based following by capitalizing on their personal brand as a *unique selling point* [6,7,41]. Furthermore, microcelebrity Instagram users consistently engaged with followers by generating discussion, which may have been in an attempt to gain social support and encourage feelings of belonging, as this has been found to be a primary motivation for Instagram use [9-11].

An interesting finding of this study was that many microcelebrity Instagram users used Instagram to share learning opportunities and describe their understanding of nursing-related clinical information. For example, microcelebrity Instagram users challenged followers to answer questions about medication administration. In addition, microcelebrity Instagram users described their perspective on the roles and responsibilities that nurses have in specific work environments, including areas such as the operating room. Sharing insights about nursing-related learning has not previously been identified as a primary reason for Instagram use. The findings from this study suggest that microcelebrity Instagram users may also use Instagram to promote educational opportunities such as courses and conferences. The reasons why microcelebrity Instagram users promoted these learning opportunities are unknown.

Humor was frequently demonstrated by microcelebrity Instagram users, often used in a manner that required the reader to have contextual knowledge to understand clinical abbreviations and terms used in the caption, in conjunction with the corresponding image. In addition, self-enhancing humor appeared to be used as a way to manage potentially stressful situations [42]. For example, microcelebrity Instagram user 08 posted a series of memes that described situations that could be

experienced in nursing. However, depending on the reader's contextual understanding, the post may be interpreted as a form of mockery for not following the directions or advice given by the health care provider.

It was also common for microcelebrity Instagram users to express their emotions through Instagram posts, which has been previously identified as a motive for Instagram use [9,11]. However, previous literature suggests that self-expression on Instagram is achieved primarily through the process of creating, editing, and posting images [9,11]. In this study, self-expression was often conveyed through heartfelt captions rather than the images themselves. Images included in the posts with meaningful captions frequently appeared to be unrelated to the caption. Most often, the unrelated images were of the microcelebrity Instagram user standing alone and smiling. In these posts, it appears as though microcelebrity Instagram users were primarily using the captions to express themselves, as opposed to creating unique images as a form of self-expression.

Five posts generated by microcelebrity Instagram users appeared to potentially depict patients, although the sources of the images and videos are unknown. As such, it is unclear if the images and videos contained patients or actors. The overwhelming consensus from nursing practice regulators is that nurses should refrain from posting any content on social media that contains patients or patient information [28,37-40]. Similarly, previous research advises nurses to refrain from posting patient information on social media [14,16-18,20,28,37-40]. However, an image posted by microcelebrity Instagram user 03 showed an open abdomen during a surgical procedure. Similarly, microcelebrity Instagram user 04 posted a series of videos showcasing an operating room environment, with what appeared to be patients awaiting surgical procedures. The uncertainty as to whether these posts show real patients or standardized patient actors demonstrates the need for further research to guide education and policy development. If the posts are depicting actual patients, there needs to be further research regarding the response of other clinicians witnessing this behavior as well as regarding how infection control practices may be compromised by staff taking pictures amid surgical procedures.

Finally, several posts showed the microcelebrity Instagram user posing in a workplace setting, dressed in work attire, with no overt patient health information displayed. At present, professional nursing guidelines do not specify whether or not it is appropriate to post an image of oneself in a workplace [37-40]. However, the Registered Nurses Association of Ontario (a professional association for Ontario nurses, who often publish clinical practice guidelines) [40] suggests that if there is uncertainty if a post is appropriate, the user should refrain from posting. Given that multiple microcelebrity Instagram users post workplace-based images, further work should be completed to determine how health care organizations interpret the depiction of their facilities via social media channels, especially when these images are instantaneously amplified by the significant following possessed by microcelebrity Instagram users.

Nurses' Social Media Use

Over the last decade, practical recommendations and policies developed by both researchers and regulatory bodies have begun to emerge in an effort to help mitigate social media–related risks by expressing the expectations of nurses and reinforcing best practice guidelines [14,17,19,28,37-40]. Much of the available advice regarding nurses' social media use portrays social media use in a negative manner, focusing on what not to do [21]. However, some literature does exist on how nurses can use social media in a positive manner. For example, nurses are encouraged to build a personal brand and express their opinions in health-related conversations on social media platforms, provided they are aware of and comply with privacy as well as relevant workplace policies [19,43]. In this study, multiple microcelebrity Instagram users developed personal brands and voiced their opinions on various nursing-related topics via Instagram. The findings of this study also demonstrate that nurses may be violating regulatory policies, as evidenced by images and videos of what appear to be patients. The findings of this study demonstrate a need for further guideline development that focuses on what nurses should be doing on social media to maintain professionalism and protect patients' rights.

Implications for Nursing Education and Practice Policies

The findings of this study highlight the need for further education and practice policy development with regard to nurses' Instagram use. Incongruencies currently exist between (1) how recommendations for integrating social media into nursing practice are enacted by some nurses; (2) the current guidelines for social media use by nurses; and (3) the ways in which nurses use social media, especially those possessing Web-based microcelebrity status. Although previous research has recommended that social media be integrated with nursing practice, current policies strongly urge nurses to avoid any and all contact with patients over social media [37-40]. In contrast, the findings of this study demonstrated that nurse microcelebrities actively engage and interact with thousands of Instagram users and generate personalized interpretations of the nursing profession via social media. In addition, in this study, multiple microcelebrity Instagram users posted images or videos that appeared to have depicted patients receiving care. Although the sources of the posts are not specified, the content of these posts directly contrasts existing guidelines for nurses' social media use. For the nursing profession, it is imperative for nursing education, professional regulatory bodies, and employers to develop more robust and dynamic policies and guidelines related to the appropriate use of social media within the profession [16,21], especially with the growing presence of Web-based nurse microcelebrities.

At present, a gap exists in the policy regarding Instagram and other social media use by nurses who have attained a microcelebrity status. The findings of this study suggest that nurses who have achieved microcelebrity status may use Instagram for a variety of purposes, and there needs to be clarity in what is considered appropriate professional behavior. Given the rapid evolution of social media, it is imperative that both

education and policy initiatives make efforts to maintain relevancy as related to new usages of these kinds of technology, including nurses who actively use these platforms to seek Web-based microcelebrity status.

Limitations

A limitation of the study was the noninteractive nature of the research, whereby only secondary analysis of preexisting Instagram photos and textual content were analyzed. Owing to the methodology and ethical requirements of the study, no participants examined in this study were contacted to comment on the findings emerging from the data analysis. Thus, ensuring appropriate representation of data, as opposed to the researcher's interpretation of data, was a challenge of this qualitative content analysis [33].

Finally, this study focused exclusively on registered nurses who had achieved a level of microcelebrity status only on Instagram. Although Instagram is one of the most popular social media platforms in 2019, different or more nuanced findings may have emerged if nursing microcelebrities who used other social media platforms (ie, YouTube and Twitter) were included in the study [1].

Future Directions

The findings of this study support the need for further research in several domains related to nurses who have attained microcelebrity status on Instagram. Further research should be completed in both the qualitative and quantitative paradigms to develop a well-rounded understanding of the concept of nurses seeking microcelebrity status on social media. Future explorations should also be conducted to determine who are following nurse microcelebrity Instagram user accounts and their reasons for doing so. The examination of merchandise endorsement as well as the reactions of microcelebrity Instagram users' followers to Instagram advertisements is also a worthwhile area for subsequent exploration. Similar research may be conducted on alternate groups of health care professionals, such as physicians, to identify if similar behaviors occur outside of the nursing profession. Finally, there is an urgent need for further research on the risks associated with nurses posting images and videos depicting patient care on Instagram.

Conclusions

The findings of this study suggest that posting characteristics and behaviors of registered nurses who have attained microcelebrity status on Instagram include (1) *engaging Instagram users*, (2) *educational opportunities and insights*, (3) *nursing-related humor*, (4) *emotions experienced by nurses*, and (5) *media and narratives including patient details or work context.*

Implications exist for nursing professionals in terms of potential privacy, professionalism, and ethical challenges associated with social media use. As social media platforms continue to thrive and evolve, nurses must be able to effectively use social media while maintaining professionalism. Nursing practice policies and guidelines must be updated to include recommendations pertinent to nurses with microcelebrity status, to maintain nurses' professionalism and to protect patients' safety.

Acknowledgments

HK's graduate studies were supported by the Ontario Graduate Scholarship (2017-2018). The authors would also like to thank the reviewers for their thoughtful and constructive feedback.

Authors' Contributions

HK and RB contributed to the design of the study. HK implemented the study design, data collection, and data analysis under the advice and supervision of RB. KJ contributed to the finalization of the analysis and generation of the manuscript.

Conflicts of Interest

None declared.

References

1. Constine J. TechCrunch. 2018 Jun 20. Instagram Hits 1 Billion Monthly Users, Up From 800 Million in September URL: https://techcrunch.com/2018/06/20/instagram-1-billion-users/ [accessed 2018-07-12]
2. Jackson D. Sprout Social: Social Media Management Solutions. 2017. Know Your Limit: The Ideal Length of Every Social Media Post URL: https://sproutsocial.com/insights/social-media-character-counter/ [accessed 2017-10-20]
3. Instagram Help Center. 2019. URL: https://help.instagram.com/ [accessed 2017-10-20]
4. Djafarova E, Trofimenko O. 'Instafamous' – credibility and self-presentation of micro-celebrities on social media. Inf Commun Soc 2019;22(10):1432-1446. [doi: 10.1080/1369118X.2018.1438491]
5. Hearn A, Schoenhoff S. From celebrity to influencer: tracing the diffusion of celebrity value across the data stream. In: Marshall PD, Redmond S, editors. A Companion to Celebrity. Hoboken, New Jersey: Wiley-Blackwell; 2015:194-211.
6. Khamis S, Ang L, Welling R. Self-branding, 'micro-celebrity' and the rise of social media influencers. Celebr Stud 2017;8(2):191-208. [doi: 10.1080/19392397.2016.1218292]
7. Marwick AE. Instafame: luxury selfies in the attention economy. Public Cult 2015 Jan 22;27(1(75)):137-160. [doi: 10.1215/08992363-2798379]

8. Marwick A, boyd D. To see and be seen: celebrity practice on Twitter. Convergence 2011 May 19;17(2):139-158. [doi: 10.1177/1354856510394539]
9. Lee E, Lee J, Moon JH, Sung Y. Pictures speak louder than words: motivations for using Instagram. Cyberpsychol Behav Soc Netw 2015 Sep;18(9):552-556. [doi: 10.1089/cyber.2015.0157] [Medline: 26348817]
10. McCain JL, Borg ZG, Rothenberg AH, Churillo KM, Weiler P, Campbell WK. Personality and selfies: narcissism and the dark triad. Comput Human Behav 2016 Nov;64:126-133. [doi: 10.1016/j.chb.2016.06.050]
11. Sheldon P, Newman M. Instagram and american teens: understanding motives for its use and relationship to excessive reassurance-seeking and interpersonal rejection. J Soc Media Soc 2019;8(1):1-16 [FREE Full text]
12. Sheldon P, Bryant K. Instagram: motives for its use and relationship to narcissism and contextual age. Comput Human Behav 2016 May;58:89-97. [doi: 10.1016/j.chb.2015.12.059]
13. Sung Y, Lee JA, Kim E, Choi SM. Why we post selfies: understanding motivations for posting pictures of oneself. Pers Individ Dif 2016 Jul;97:260-265. [doi: 10.1016/j.paid.2016.03.032]
14. Alexander S. Polishing your presence online: wise use of social media to enhance your reputation. Clin Nurse Spec 2016;30(5):261-263. [doi: 10.1097/NUR.0000000000000238] [Medline: 27509561]
15. Caulfield T, Marcon AR, Murdoch B, Brown JM, Perrault ST, Anthony SJ, et al. Health misinformation and the power of narrative messaging in the public sphere. Can J Bioeth 2019;2(2):52-60. [doi: 10.7202/1060911ar]
16. Cleary M, Ferguson C, Jackson D, Watson R. Editorial: social media and the new e-professionalism. Contemp Nurse 2013 Oct;45(2):152-154. [Medline: 24422224]
17. National Council of State Boards of Nursing. 2011. White Paper: A Nurse's Guide to the Use of Social Media URL: https://www.ncsbn.org/Social_Media.pdf [accessed 2019-12-14]
18. Balestra M. American Nurse Today. 2018 Mar 6. Social Media Missteps Could Put Your Nursing License at Risk URL: https://www.americannursetoday.com/social-media-nursing-license-risk/ [accessed 2019-12-14]
19. Reinbeck D, Antonacci J. How nurses can use social media to their advantage. Nursing 2019 May;49(5):61-63. [doi: 10.1097/01.NURSE.0000554624.05347.6e] [Medline: 31022033]
20. Miller LA. Social media savvy: risk versus benefit. J Perinat Neonatal Nurs 2018;32(3):206-208. [doi: 10.1097/JPN.0000000000000346] [Medline: 30036302]
21. Ryan G. International perspectives on social media guidance for nurses: a content analysis. Nurs Manag (Harrow) 2016 Dec 1;23(8):28-35. [doi: 10.7748/nm.2016.e1555] [Medline: 27905234]
22. Ferguson C. It's time for the nursing profession to leverage social media. J Adv Nurs 2013 Apr;69(4):745-747. [doi: 10.1111/jan.12036] [Medline: 23488814]
23. Green J. Nurses' online behaviour: lessons for the nursing profession. Contemp Nurse 2017 Jun;53(3):355-367. [doi: 10.1080/10376178.2017.1281749] [Medline: 28077035]
24. Hazzam J, Lahrech A. Health care professionals' social media behavior and the underlying factors of social media adoption and use: quantitative study. J Med Internet Res 2018 Nov 7;20(11):e12035 [FREE Full text] [doi: 10.2196/12035] [Medline: 30404773]
25. Kotsenas AL, Arce M, Aase L, Timimi FK, Young C, Wald JT. The strategic imperative for the use of social media in health care. J Am Coll Radiol 2018 Jan;15(1 Pt B):155-161. [doi: 10.1016/j.jacr.2017.09.027] [Medline: 29128501]
26. HIPAA Journal - News and articles about HIPAA. 2018. Texas Nurse Fired for Social Media HIPAA Violation URL: https://www.hipaajournal.com/texas-nurse-fired-for-social-media-hipaa-violation/ [accessed 2019-12-14]
27. Ornstein C. ProPublica. 2015 Dec 21. Inappropriate Social Media Posts by Nursing Home Workers, Detailed URL: https://www.propublica.org/article/inappropriate-social-media-posts-by-nursing-home-workers-detailed [accessed 2019-12-14]
28. National Council of State Boards of Nursing. 2018. A Nurse's Guide to the Use of Social Media URL: https://www.ncsbn.org/NCSBN_SocialMedia.pdf [accessed 2019-12-14]
29. Wang Z, Wang S, Zhang Y, Jiang X. Social media usage and online professionalism among registered nurses: a cross-sectional survey. Int J Nurs Stud 2019 Oct;98:19-26. [doi: 10.1016/j.ijnurstu.2019.06.001] [Medline: 31255853]
30. Kress D, Godack CA, Berwanger TL, Davidson PM. The new script of nursing: using social media and advances in communication - to create a contemporary image of nursing. Contemp Nurse 2018;54(4-5):388-394. [doi: 10.1080/10376178.2018.1537720] [Medline: 30474519]
31. Lupton D, Michael M. 'For me, the biggest benefit is being ahead of the game': the use of social media in health work. Soc Media Soc 2017 Apr 11;3(2):205630511770254. [doi: 10.1177/2056305117702541]
32. Sandlin JK, Hinmon D. Beyond baby steps: today's use of social networking sites and the nursing profession. J Perinat Neonatal Nurs 2016;30(3):204-208. [doi: 10.1097/JPN.0000000000000189] [Medline: 27465450]
33. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
34. Booth RG. Happiness, stress, a bit of vulgarity, and lots of discursive conversation: a pilot study examining nursing students' tweets about nursing education posted to Twitter. Nurse Educ Today 2015 Feb;35(2):322-327. [doi: 10.1016/j.nedt.2014.10.012] [Medline: 25467718]

35. Kostick KM, Blumenthal-Barby JS, Wilhelms LA, Delgado ED, Bruce CR. Content analysis of social media related to left ventricular assist devices. Circ Cardiovasc Qual Outcomes 2015 Sep;8(5):517-523. [doi: 10.1161/CIRCOUTCOMES.115.002032] [Medline: 26219889]
36. Snelson C. Vlogging about school on YouTube: An exploratory study. New Media Soc 2015;17(3):321-339. [doi: 10.1177/1461444813504271]
37. American Nurses Association: ANA Enterprise. Social Media URL: https://www.nursingworld.org/social/ [accessed 2019-12-14]
38. International Nurse Regulator Collaborative. College of Nurses of Ontario. 2016. Social Media Use: Common Expectations for Nurses URL: http://www.cno.org/globalassets/docs/prac/incr-social-media-use-common-expectations-for-nurses.pdf [accessed 2017-11-15]
39. Ontario Nurses Association. 2016. How to Navigate Social Media Effectively: A Best Practices Primer From Your Union URL: https://www.ona.org/wp-content/uploads/ona_socialmediabestpracticesf_201711.pdf [accessed 2017-11-15]
40. Registered Nurses Association of Ontario. 2015. Use of Social Media URL: https://rnao.ca/sites/rnao-ca/files/Section_7_Use_of_Social_Media.pdf [accessed 2017-11-15]
41. Whitmer JM. You are your brand: self-branding and the marketization of self. Sociol Compass 2019 Jan 21;13(3):e12662. [doi: 10.1111/soc4.12662]
42. Reyes A, Rosso P, Buscaldi D. From humor recognition to irony detection: the figurative language of social media. Data Knowl Eng 2012;74:1-12. [doi: 10.1016/j.datak.2012.02.005]
43. Shattell M, Darmoc R. Becoming a public thought leader in 140 characters or less: how nurses can use social media as a platform. J Psychosoc Nurs Ment Health Serv 2017 Jun 1;55(6):3-4. [doi: 10.3928/02793695-20170519-06] [Medline: 28585660]

Abbreviations

HIPAA: Health Insurance Portability and Accountability Act

Edited by E Borycki; submitted 17.10.19; peer-reviewed by D Dordunoo, J Hazzam; comments to author 30.11.19; revised version received 04.02.20; accepted 11.03.20; published 26.05.20.

Please cite as:
Kerr H, Booth R, Jackson K
Exploring the Characteristics and Behaviors of Nurses Who Have Attained Microcelebrity Status on Instagram: Content Analysis
J Med Internet Res 2020;22(5):e16540
URL: http://www.jmir.org/2020/5/e16540/
doi:10.2196/16540
PMID:32452809

Original Paper

The Mediator Roles of Problematic Internet Use and Perceived Stress Between Health Behaviors and Work-Life Balance Among Internet Users in Germany and China: Web-Based Cross-Sectional Study

Lingling Gao[1], MSc; Yiqun Gan[2], PhD; Amanda Whittal[1], PhD; Song Yan[1], PhD; Sonia Lippke[1], PhD

[1]Department of Psychology & Methods, Jacobs University Bremen, Bremen, Germany

[2]School of Psychological Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, China

Corresponding Author:
Yiqun Gan, PhD
School of Psychological Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health
Peking University
5 Yiheyuan Rd, Haidian District
Beijing
China
Phone: 86 10 62757271
Email: ygan@pku.edu.cn

Abstract

Background: Work-life balance is associated with health behaviors. In the face of digitalization, understanding this link requires a theory-based investigation of problematic internet use and perceived stress, which are so far unknown.

Objective: On the basis of the compensatory carry-over action model, this study aimed to determine whether problematic internet use and perceived stress mediate the relationship between health behaviors and work-life balance in two groups of internet users from different environments (residents in Germany and China). We also investigated whether the place of residence was a moderator.

Methods: An online questionnaire (N=877) was administered to residents from Germany (n=374) and China (n=503) in 3 languages (German, English, and Chinese). Moderated mediation analyses were run with health behaviors as the independent variable, work-life balance as the dependent variable, problematic internet use and perceived stress as the mediator variables, and place of residence as a potential moderator.

Results: On a mean level, individuals in Germany reported less problematic internet use and more health behaviors than individuals in China; however, they also had lower work-life balance and higher perceived stress. Results showed that health behaviors seem to be directly related to work-life balance in both groups. Among the residents of Germany, a partial mediation was revealed (β=.13; *P*=.01), whereas among the residents of China, a full mediation was found (β=.02; *P*=.61). The mediator role of perceived stress was compared with problematic internet use in all the serial models and the parallel model. Residence moderated the relationship between health behaviors and work-life balance: The interrelation between health behaviors and work-life balance was stronger in Germany (β=.19; *P*<.001) than in China (β=.11; *P*=.01) when controlling for other variables.

Conclusions: The findings of this study are in line with the compensatory carry-over action model. To promote work-life balance, individuals should perform health behaviors to help overcome problematic internet use and perceived stress. Both problematic internet use and perceived stress mediated health behaviors and work-life balance partially in German study participants and fully in Chinese study participants.

(J Med Internet Res 2020;22(5):e16468) doi:10.2196/16468

KEYWORDS

healthy lifestyle; work-life balance; internet; healthy diet; exercise; culture

Introduction

Work-Life Balance

Increasingly, many people struggle with maintaining a work-life balance [1] and have to manage issues arising from a poor work-life balance, such as sickness, limitations in the ability to work, and absence of work [2]. There has been increasing attention on how best to achieve a work-life balance in the modern society, which has been investigated by researchers from different perspectives. Work time (eg, long work hours) has been found to have a negative association with work-life balance [3-5]. When exploring the ways to decrease this issue, it was found that the internet can help change the traditional work model to a more flexible model. The use of emails, smartphones, and home computers, for example, makes it possible to work without having a typical *nine-to-five* workday [6], which may help to maintain a work-life balance. However, it was also found that excessive internet use can decrease work-life balance [7-9]. Perceived stress was also found to be an important factor that affects work-life balance negatively [10], but which could be mitigated by health behaviors [9]. Previous evidence has indicated that health behaviors (eg, regular physical activity and healthy diet) help to prevent stress [11]. Moreover, health behaviors, such as regular physical activity and healthy nutrition of employees, have also been recommended to maintain a work-life balance [12-14].

Health Behaviors

Common important health behaviors include both health-promoting behaviors [15], for instance, physical activity [16,17] and a healthy diet (eg, sufficient water intake and fruit and vegetable intake) [18,19], and health-risk behaviors [16,17], such as smoking status [16,20,21] and alcohol consumption status [22,23]. Previous studies have found that different single health behaviors are interrelated [16], such as diet and exercise [24,25] and smoking and alcohol consumption [18]. There is increasing focus on how to help employees maintain their work-life balance. Besides efforts by employers, employees can also help themselves to improve their work-life balance by performing health behaviors after work. Employees' health behaviors are the focus of this study that sought to explore how people's health behaviors interrelate with work-life balance.

Perceived Stress and Problematic Internet Use as Mediators

As described earlier, not only health behaviors but also problematic internet use and perceived stress have been found to play a role in work-life balance [10,26]. The *compensatory carry-over action model* by Lippke [27] provides an approach to understanding the underlying psychological and behavioral mechanisms of multiple behaviors related to an outcome. However, the model has not yet been applied to the main focus of this paper, that is, work-life balance, internet use, and health behaviors. According to this model, different behaviors are not isolated; rather, they interact with each other in complex ways. Emotionally relevant higher-level goals (such as work-life balance) drive different health behaviors by initiating and strengthening behavior-specific intentions [27]. If individuals have set a higher-level goal and did not reach this goal (eg, because of situations such as work stress), their well-being is affected [28-30]. As previous studies have indicated, most people experience different stressors within their lives, which can interfere with adopting and maintaining behaviors [27]. In line with this model, perceived stress has been found to be a negative indicator of quality of life [31-33]. Moreover, it has been found that stress acts as a mediator between work-family conflict and psychological health [34]. Previous evidence has also indicated that health behaviors (eg, regular physical activity and healthy diet) help to prevent stress [11] and maintain a work-life balance [12-14]. Among a sample of working individuals, higher perceived stress was found to be associated with a higher fat diet and less physical activity [35,36]. Perceived stress may be one factor that undermines the positive relationship between health behaviors and work-life balance. However, the mediator role of perceived stress between health behaviors and work-life balance has received limited attention in previous research. This study is a novel effort to fill this gap, in particular, to help understand the interrelation of different health behaviors and internet use behavior on the basis of the *compensatory carry-over action model*.

Previous studies have also found that poor health behaviors, such as inadequate nutrition intake, were related to problematic internet use [7-9]. According to the *compensatory carry-over action model* [27], the outcome experience (eg, problematic internet use) from one behavior (eg, internet use behavior) may also affect a higher-level goal (eg, work-life balance) together with another behavior (eg, health behavior). Although the internet may help to arrange flexible work time and in turn maintain work-life balance [6], for some people, problematic internet use can also harbor hazards to their work-life balance [7-9]. Moreover, time is a finite resource, the use of which must be divided among individual needs, work, and family needs [37]. Problematic internet use may occupy the time one could spend on physical activity, family, and work and is thus negatively related to work-life balance. Excessive internet use has been found to be associated with behaviors that negatively affect people's health, mood, work, and occupation [38]. So far, little is known about the relationship between health behaviors and work-life balance, while also considering the risk brought by problematic internet use. Owing to the large number of people suffering from the consequences of problematic internet use [39], it is important to examine how health behaviors combined with problematic internet use interrelate with work-life balance.

Some studies have indicated that problematic internet use has a direct impact on perceived stress [40,41], for example, it was found that problematic internet use is related to perceived stress and negative feelings, such as depression and anxiety [42,43]. Other studies have shown that perceived stress plays an important role in problematic internet use [44,45]. According to the *compensatory carry-over action model* [27], both perceived stress and problematic internet use may mediate the relationship between health behaviors and a higher-level goal (ie, work-life balance).

In this study, this underlying mechanism of this relationship was explored by comparing two possible serial models, with perceived stress and problematic internet use as serial mediators.

If problematic internet use is assumed to influence perceived stress, then the chain is as follows: health behavior → problematic internet use → perceived stress → work-life balance. If one assumes that perceived stress influences problematic internet use, then the chain is as follows: health behavior → perceived stress → problematic internet use → work-life balance.

Place of Residence Differences Among Study Variables

Previous studies have shown large cultural differences between China and Germany [46,47]. For example, according to Hofstede's empirical framework, Germany is representative of a *Western* country and individualistic society, in which people consider their needs over the needs of the organization to which they belong. Conversely, China is representative of an *Eastern* country and collectivistic society, in which people put the needs of the organization before individual needs [48-52].

On the basis of these differences, previous studies comparing work-life balance between *Western* and *Eastern* companies revealed that work-life balance ranks higher in Western companies than in Eastern companies, as more attention is given to it [53]. The contexts of work-life conflict and solution in China are significantly different from those found in Western countries [54]. Some studies have found that in Eastern countries, family, including elderly parents, may provide advice and emotional support when people need help in maintaining a work-life balance [55,56]. Culture has been found to moderate the relationship between work resources and work-family conflict [57]. Although health behaviors as personal resources [58] were found to be associated with work-life balance [12-14], the moderator role of the place of residence between health behaviors and work-life balance still needs to be investigated.

The difference in health behaviors between Western countries, such as European countries and the United States, and Eastern countries, such as China, was investigated in previous studies [59,60]. Factors such as the geographic, historical, and economic development stage of a country and the coping mechanisms of its citizens may affect people's health behaviors, such as food preferences and physical activity [61,62]. For example, the number of facilities for physical activity provided by the government (eg, parks) may inhibit or facilitate participation in physical activity, depending on the socioeconomic status [62]. It has been found that fruit intake and physical activity have higher prevalence in Western countries [63], and the percentage of nonsmokers is higher in Eastern countries than that in Western counties [63]. Differences in perceived stress exist between Western and Eastern countries as well [64,65]. For problematic internet use, previous studies have shown that Eastern counties have a higher internet addiction score than Western countries [66,67].

Proposed Research Questions

On the basis of these previous research findings and the *compensatory carry-over action model*, this study sought to understand and explore the mechanisms underlying the relationship between health behaviors and work-life balance.

The following two major research questions were addressed:

1. Do problematic internet use and perceived stress mediate the relationship between health behaviors and work-life balance in both Germany and China?
2. Does the country of residence moderate the relationship between a healthy lifestyle and work-life balance?

Methods

Participants

The online survey included 877 participants (538/877, 61.3% women). German (98/877, 11.2%) and English (276/877, 31.5%) versions of the questionnaire were used to collect data in Germany (group 1: residents in Germany), and a Chinese (503/877, 57.4%) version of the questionnaire was used to collect data in China (group 2: residents in China). A minimum sample size of 233 was used in both groups. The participants' ages ranged from 17 to 65 years (mean 30.0, SD 10.8).

Procedure

Data were gathered from residents in middle-sized cities in Germany (Bremen) and China (Shijiazhuang). As Germany is the second most popular migration destination in the world [68] and English is widely spoken in Germany, both German and English versions of the questionnaire were provided for the participants to freely choose the language they prefer (no significant differences were found between the German and English versions of the questionnaire in the pilot analyses). Residents in China were provided with the Chinese version of the questionnaire.

The online questionnaires were used to collect data from October 2016 to August 2018 through email (in Germany and China), Facebook (in Germany), and link distribution face to face (with the help of the research assistants in Germany and with the help of primary school teachers who distributed the questionnaires among the pupils' parents at parent-teacher conferences in China). Participants answered multiple-choice questions by clicking the appropriate check box and open questions by inputting text content. On the first page of the survey, all participants were informed of the confidentiality, anonymity, and voluntary nature of their responses and given the opportunity to obtain the study results by providing contact information on the page. Those who clicked the box to provide their informed consent could continue to the questionnaire pages. The study received ethical approval by the Ethics Commission of the German Association of Psychology (Deutsche Gesellschaft für Psychologie, EK-A-SL022013). The link for the survey was sent to the staff and students in universities. Study participants were also asked to forward the invitation and the link to families and friends in Germany and China.

Measures

Health behavior was operationalized from the summation of 5 items: physical activity, sufficient water consumption, nutrition, smoking status, and alcohol consumption status. For the purposes of analysis, each behavior was coded into a binary variable, with 0 representing the unhealthy option. *Physical activity* was assessed with the short form of the International Physical Activity Questionnaire [69], which has been tested for use in adults and has shown cross-cultural validity and reliability

[70]. This questionnaire covers mild, moderate, and strenuous physical activities. Participants indicated the duration (in minutes) and frequency (in days) of each activity domain per week. Minutes per week were multiplied by days per week to obtain a sum score per activity domain and then added for total activity per week. Total physical activity was classified as 1 (those who fulfilled the World Health Organization's recommendation of ≥2.5 hours per week) and 0 (physical activity <2.5 hours per week).

S*ufficient water consumption* was assessed by the question, "Please think about your typical weeks, do you drink 1.5 L of non-alcoholic and non-caffeinated beverage (water, juice, fruit and herbal tea) during the day?" and it was possible to answer with yes (1) or no (0). This question has been shown to have acceptable reliability and validity [16]. *Nutrition* was measured by the question, "Please think about your typical weeks, do you eat five or more servings of fruit and vegetable per day?" and was to be answered with yes or no. This question has also been shown to have acceptable reliability and validity [71]. *Smoking status* was assessed by asking participants "Are you a smoker?" Answers were classified into 1 (nonsmoker or ex-smoker) and 0 (occasional smoker or regular smoker). The *alcohol consumption status* was assessed by asking participants "Do you drink alcohol on a regular basis?" and was classified as 1 (no) and 0 (yes). Both the smoking and alcohol questions have been shown to have acceptable reliability and validity [72]. All different health behaviors were summed, with 0 indicating the fewest health behaviors performed and 5 indicating the performance of all health behaviors on the recommended level.

Work-life balance was measured by using a 5-item scale [73] consisting of employees' satisfaction with their achieved balance between work and private life. Answers were on a 5-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree). Sample items were "I am satisfied with the balance between my work and private life."; "It is difficult for me to balance my work and private life"; and "I am meeting the requirements of both my work and my private life." The Cronbach alpha was .81 in this study.

Perceived stress was measured with 2 questions from the Perceived Stress Scale [74], as in previous studies [75]—"In the last month, how often have you felt nervous and stressed?" and "In the last month, how often have you felt difficulties were piling up so high that you could not overcome them?"—and it was rated on a 5-point Likert scale. The Cronbach alpha was .75 in this study.

Problematic internet use was assessed with the Internet Addiction Questionnaire (ISS-10r) [76]. The current version includes 10 items with answer scores on a scale of 1 to 4 ranging from completely disagree to agree completely (eg, "I often spend more time on the internet than I intended"). For the evaluation of problematic internet use, ISS-10r sum values of the 5 subscales (loss of control, withdrawal symptoms, development of tolerance, negative consequences on work and performance, and negative consequences on social relationships) were utilized. The total score can vary from 20 to 80 and distinguishes 3 types of internet use: no problematic internet use (20-49), problematic internet use tendency (50-59), problematic internet use (60-80). The ISS-10r is one of the validated tools to assess problematic internet use, and the Cronbach alpha was .81 in this study.

Participants also answered sociodemographic questions, including those on age, gender, height, weight, marital status, employment status, and education level.

Statistical Analysis

A correlation analysis was performed to investigate the relationships between health behaviors, perceived stress, problematic internet use, and work-life balance in both groups (group 1: residents in Germany and group 2: residents in China). Moderated mediation analyses were run with health behaviors as the independent variable, work-life balance as the dependent variable, problematic internet use and perceived stress as the mediator variables, and place of residence as a moderator in both groups. Moderated mediation analyses were performed using SPSS version 24 (IBM Corp) and PROCESS [77]. Significance was accepted at an alpha level of .05.

Results

Preliminary Analyses

In this survey, 42.6% (374/877) of participants completed the German or English version of the questionnaire (group 1: residents in Germany), and 57.4% (503/877) of participants completed the Chinese version of the questionnaire (group 2: residents in China). When comparing health behavior items between these 2 different places of residence groups, *Physical activity* participation was higher in group 1 than in group 2 (245/374, 65.5% vs 277/503, 55.1%; *P*=.002), more participants have S*ufficient water consumption* in group 1 than in group 2 (270/374, 72.2% vs 213/503, 42.3%; *P*<.001), more participants drink alcohol on a regular basis (*alcohol consumption status*) in group 1 than in group 2 (180/374, 48.1% vs 125/503, 24.9%; *P*<.001), and there were no significant differences in *Nutrition* (*P*=.35) and *Smoking status* (*P*=.85) between the 2 groups. When comparing other variables between these 2 different places of residence groups, problematic internet use and work-life balance scores of group 2 were significantly higher than those of group 1, whereas perceived stress and health behaviors were lower in group 2. Descriptive statistics are presented in Table 1.

Table 1. Descriptive statistics of main study variables in the 2 groups (N=877; group 1: residents in Germany and group 2: residents in China).

Variables	Group 1, mean (SD)	Group 2, mean (SD)	Total		*t* test (df)	*P* value
			Mean (SD)	Range		
Age (years)	26.7 (12.5)	32.5 (8.4)	30.0 (10.8)	17-65	−8.24 (875)	<.001
Problematic internet use	41.8 (11.6)	44.2 (11.9)	43.2 (11.8)	20-80	−2.96 (875)	.003
Perceived stress	5.3 (1.8)	4.6 (1.5)	4.9 (1.6)	2-10	6.32 (875)	<.001
Health behaviors	3.0 (1.1)	2.9 (1.2)	2.9 (1.2)	0-5	2.44 (875)	.02
Work-life balance	16.5 (4.5)	18.0 (4.1)	17.4 (4.3)	5-25	−5.36 (875)	<.001

Correlation Analyses in Different Groups

Table 2 displays the bivariate Pearson correlation between health behaviors, problematic internet use, perceived stress, and work-life balance. In both groups, health behaviors were positively correlated with work-life balance, whereas problematic internet use and perceived stress were negatively related to a healthy lifestyle and work-life balance.

Table 2. Pearson correlation of variables in 2 groups.

Group 1[a]	Group 2[b]					
	1	2	3	4	5	6
1. Gender[c]	N/A[c]	−0.07	−0.19	0.10	0.02	−0.04
2. Age	−0.11	N/A	−0.05	−0.36	−0.20	0.15
3. Health behaviors	−0.04	−0.06	N/A	−0.12	−0.27	0.11
4. Problematic internet use	0.15	−0.40	−0.16	N/A	0.32	−0.22
5. Perceived stress	−0.02	−0.23	−0.14	0.35	N/A	−0.32
6. Work-life balance	0.01	0.16	0.19	−0.27	−0.45	N/A

[a]Group 1: residents in Germany; correlations presented below the diagonal.

[b]Goup 2: residents in China; correlations presented above the diagonal.

[c]Gender was dummy coded such that 1=female and 2=male. $r \geq .15$, $P=.01$; $0.10 \leq r \leq .14$, $P=.05$.

[d]N/A: not applicable.

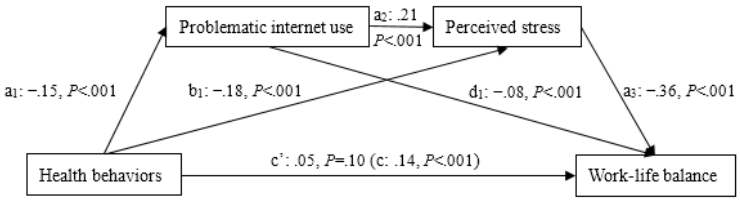
Testing for Mediation Effects

The first aim of the study was to examine whether problematic internet use and perceived stress mediate the relationship between a healthy lifestyle and work-life balance in general. Two possible serial mediation models and one parallel model were estimated and assessed to investigate the mediator roles of problematic internet use and perceived stress.

Serial Mediation

Serial mediation assumes "a causal chain linking the mediators, with a specified direction of causal flow" [78]. Model 6 of PROCESS was used to test 2 possible chains: (1) health behavior → problematic internet use → perceived stress → work-life balance and (2) health behavior → perceived stress → problematic internet use → work-life balance. The proposed research model for chain 1 and results after controlling for covariates (age and gender) are presented in Figure 1. The mediation effect accounted for 19% of the total effect: $F_{5,871}=39.78$; $P<.001$. The residual direct effect was not significant (c' in Figure 1). There were significant indirect paths from health behaviors to work-life balance through problematic internet use (a_1d_1, β=.01, 95% CI 0.002 to 0.03) and perceived stress (b_1a_3, β=.06, 95% CI 0.04 to 0.09). Furthermore, the indirect path from health behaviors to work-life balance through both problematic internet use and perceived stress was also significant ($a_1a_2a_3$, β=.01, 95% CI 0.01 to 0.02). When substituting the order of the mediators to test chain 2, the indirect path remained significant with $P<.001$, but the effect was weaker (β=.003, 95% CI 0.001 to 0.01).

Figure 1. Serial mediation model and path coefficients predicting work-life balance. Path $a_1 \rightarrow a_2 \rightarrow a_3$ is the full serial mediation path (1). Path a_1d_1 is the path from health behavior to work-life balance through problematic internet use. Path b_1a_3 is the path from health behavior to work-life balance through perceived stress.



Parallel Mediation

Parallel mediation assumes that problematic internet use and perceived stress do not influence each other when they mediate the relationship between health behaviors and work-life balance. Model 4 of PROCESS [77] was used to examine parallel mediation. The results of all participants are displayed in Table 3. After controlling for covariates (ie, age and gender), the results revealed that health behaviors were negatively associated with problematic internet use (model 2), which in turn was also interrelated with work-life balance (model 4). In addition, health behaviors were found to be negatively associated with perceived stress (model 3), which was also linked negatively to work-life balance (model 4).

The residual direct effect was not significant, with P=.10 (model 4). The indirect effect of problematic internet use was β=.01, 95% CI 0.002 to 0.03, and the indirect effect of perceived stress was β=.08, 95% CI 0.05 to 0.11. Most variance could be explained in model 4, predicting work-life balance with health behaviors, problematic internet use, and perceived stress.

Results exploring the mediation effects in group 1 and group 2 are shown in Figure 2. In group 1, the residual direct effect was still significant (β=.13; P=.01) after including the mediators indicating a partial mediation only. The indirect effect of problematic internet use was β=.02, 95% CI 0.003 to 0.05; the indirect effect of perceived stress was β=.07, 95% CI 0.03 to 0.11. The mediation effect accounted for 23% of the total effect ($F_{5,368}$=22.48; P<.001).

In group 2, the residual direct effect was not significant (β=.02; P=.61) when including the mediators, indicating a full mediation. The indirect effect of problematic internet use was β=.01, 95% CI 0.001 to 0.04; the indirect effect of perceived stress was β=.07, 95% CI 0.04 to 0.11. The mediation effect accounted for 12% of the total effect ($F_{5,497}$=14.06; P<.001).

Table 3. Mediation effects of health behaviors on work-life balance.

Predictors	Model 1[a] (work-life balance)		Model 2[b] (problematic internet use)		Model 3[c] (perceived stress)		Model 4[d] (work-life balance)	
	β (95% CI)	*P* value	β (95% CI)	*P* value	β (95% CI)	*P* value	β (95% CI)	*P* value
Gender[e]	.04 (−0.10 to 0.18)	.53	.14 (0.02 to 0.27)	.03	−0.10 (−0.23 to 0.03)	.14	.02 (−0.11 to 0.15)	.76
Age (years)	.02 (0.01 to 0.03)	<.001	−0.03 (−0.04 to −0.03)	<.001	−0.03 (−0.03 to −0.02)	<.001	.01 (0.001 to 0.01)	.02
Health behaviors	.14 (0.08 to 0.21)	<.001	−0.15 (−0.22 to −0.09)	<.001	−0.21 (−0.28 to −0.15)	<.001	.05 (−0.01 to 0.12)	.10
Problematic internet use	N/A[f]	N/A	N/A	N/A	N/A	N/A	−0.08 (−0.15 to −0.01)	.02
Perceived stress	N/A	N/A	N/A	N/A	N/A	N/A	−0.36 (−0.42 to −0.29)	<.001

[a]R^2=0.06, $F_{3,873}$=17.73, P<.001.

[b]R^2=0.14, $F_{3,873}$=47.55, P<.001.

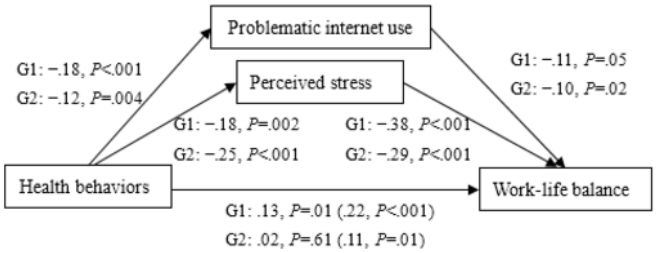
[c]R^2=0.11, $F_{3,873}$=36.83, P<.001.

[d]R^2=0.19, $F_{5,871}$=39.78, P<.001.

[e]Gender was dummy coded with 1=female and 2=male.

[f]N/A: not applicable.

Figure 2. Multiple mediation model with standardized regression coefficients predicting work-life balance in both groups. G1, Group 1: residents in Germany; G2, Group 2: residents in China.



Testing for Mediated Moderation

The second aim of the study was to examine whether place of residence moderates the relationship between health behavior and work-life balance. This was tested with a mediated moderation analysis (model 8) of PROCESS [77], and the results are displayed in Table 4. In model 1, the total effect of the health behaviors on work-life balance was significant, and this association was not moderated by the place of residence. In model 2, the direct effect of health behaviors on problematic internet use was significant, and this association was also not moderated by the place of residence. In model 3, the direct effect of health behaviors on perceived stress was significant, and this association was also not moderated by the place of residence. In model 4, the direct effect of health behaviors on work-life balance was significant, and there was a significant moderation effect of the place of residence.

To further demonstrate the pattern of this moderation effect, the interaction effect by simple slopes between health behaviors and place of residence on work-life balance is plotted in Figure 3. The simple slope test revealed that in group 1, the interrelation between health behaviors and work-life balance was significant (β=.19; *P*<.001), whereas this relation was weaker in group 2 (β=.11; *P*=.01). All predictors accounted for 20% of the variance in work-life balance in this model.

Table 4. Mediated moderation effects of health behaviors on work-life balance.

Predictors	Model 1[a] (work-life balance)		Model 2[b] (problematic internet use)		Model 3[c] (perceived stress)		Model 4[d] (work-life balance)	
	β (95% CI)	*P* value	β (95% CI)	*P* value	β (95% CI)	*P* value	β (95% CI)	*P* value
Gender[e]	.03 (−0.10 to 0.17)	.63	.14 (0.02 to 0.27)	.03	−0.10 (−0.23 to 0.03)	.13	.02 (−0.11 to 0.14)	.82
Age (years)	.02 (0.01 to 0.02)	<.001	−0.04 (−0.04 to 0.00)	<.001	−0.02 (−0.03 to −0.02)	<.001	.01 (−0.00 to 0.01)	.15
Health behaviors	.15 (0.09 to 0.22)	<.001	−0.15 (−0.21 to −0.08)	<.001	−0.22 (−0.28 to −0.16)	<.001	.07 (0.003 to 0.13)	.04
Residence	.15 (0.08 to 0.21)	<.001	.19 (0.13 to 0.25)	<.001	−0.16 (−0.22 to −0.10)	<.001	.11 (0.05 to 0.18)	<.001
Problematic internet use	N/A[f]	N/A	N/A	N/A	N/A	N/A	−0.10 (−0.17 to −0.04)	.003
Perceived stress	N/A	N/A	N/A	N/A	N/A	N/A	−0.34 (−0.40 to −0.27)	<.001
Health behaviors × residence	−0.05 (−0.12 to 0.01)	.11	−0.04 (−0.03 to 0.10)	.25	−0.04 (−0.10 to 0.03)	.28	−0.06 (−0.12 to 0.00)	.05

[a]R^2=0.08, $F_{5,871}$=15.09, *P*<.001.

[b]R^2=0.18, $F_{5,871}$=37.15, *P*<.001.

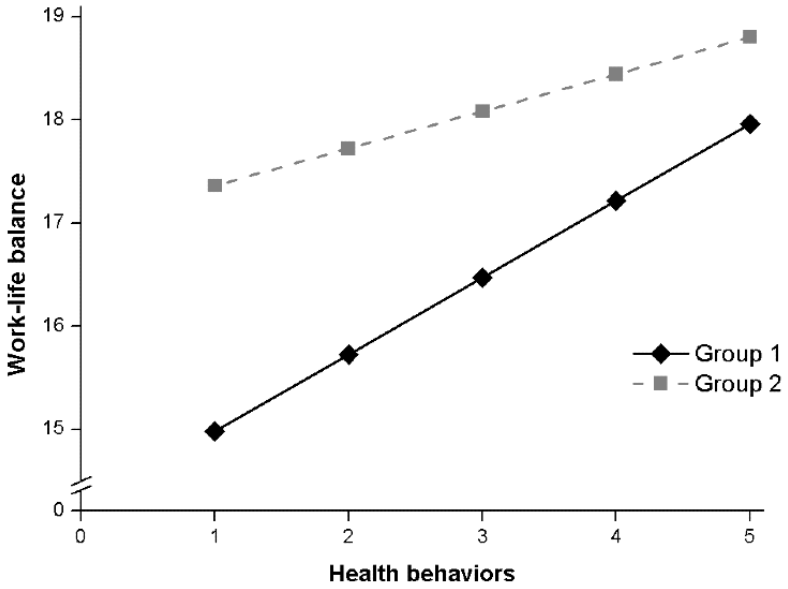
[c]R^2=0.14, $F_{5,871}$=27.97, *P*<.001.

[d]R^2=0.20, $F_{7,869}$=31.04, *P*<.001.

[e]Gender was dummy coded with 1=female and 2=male.

[f]N/A: not applicable.

Figure 3. Interaction of health behaviors and residence on work-life balance (Group 1: residents in Germany; Group 2: residents in China).



Discussion

Principal Findings

An online questionnaire assessed residents from Germany (n=374) and China (n=503) in 3 languages (German, English, and Chinese). On a mean level, individuals in Germany reported less problematic internet use and more health behaviors than individuals in China; however, they also had lower work-life balance and higher perceived stress. This study is one of the first studies to explore the potential mechanisms through which health behaviors impact work-life balance [10,26,79]. Results revealed that health behaviors were indirectly related to work-life balance through perceived stress and problematic internet use, and the mediator role of problematic internet use was not prominent in either of the serial models or the parallel model. There was a partial mediation in Germany (group 1) and a full mediation in China (group 2). The relation between health behaviors and work-life balance was significant in group 1, whereas it was weaker in group 2. This study utilized the *compensatory carry-over action model* as the underlying understanding in investigating how health behaviors interrelate with employees' work-life balance.

Mediator Roles of Stress and Problematic Internet Use

When testing the mediator roles of stress and problematic internet use, in both serial and parallel mediation models, the residual direct effect was not significant, which indicated that problematic internet use and perceived stress fully mediated the indirect association between health behaviors and work-life balance. In the serial mediation models, the chain health behavior → problematic internet use → perceived stress → work-life balance was superior to the chain health behavior → perceived stress → problematic internet use → work-life balance. However, the effect size was small. Neither of the proposed serial models revealed a large mediating effect. In the parallel model, results exploring the mediation effects in the Germany group and the China group revealed that in the Germany group, the residual direct effect was significant. This may indicate that problematic internet use and perceived stress partially mediated the indirect association between health behaviors and work-life balance. In the China group, the residual direct effect was not significant, which is because problematic internet use and perceived stress fully mediated the indirect association between health behaviors and work-life balance.

Although health behaviors in general were better in Germany than in China, the work-life balance score was lower. According to the model, this was because of the significant difference in the mediator-perceived stress, which was much higher in Germany than in China. This is in line with previous studies [79].

The mediator role of perceived stress in this study is consistent with previous studies, which found that perceived stress could be mitigated by health behaviors [9] and affected work-life balance negatively [10]. Thus, perceived stress management is of high importance. This result also supports the *compensatory carry-over action model*, as work-life balance is not only associated with predictor variables related to health behaviors but also with mediator variables, such as perceived stress. To the best of our knowledge, this is the first study that clarifies the mediator roles of perceived stress and problematic internet use in a healthy lifestyle consisting of different single health behaviors and work-life balance in different countries of residence.

Although previous studies have found health behaviors to be negatively associated with problematic internet use [7-9], which in turn affects work-life balance [80-82], in this study, the meditation role of problematic internet use was not prominent compared with the perceived stress mediator. This may be because of the interaction between perceived stress and problematic internet use. This study revealed that there was a significant relationship between problematic internet use and perceived stress. Previous studies have indicated that problematic internet use has a direct impact on perceived stress [40,41]; however, perceived stress may also come from other aspects, such as occupational, interpersonal, or financial stresses

[83]. Moreover, previous research has demonstrated that perceived stress plays an important role in problematic internet use [44,45]. It is difficult to judge which factor affects the other, as it depends on circumstances, and this study was only cross-sectional.

Apart from perceived stress and problematic internet use, the path between health behaviors and work-life balance may also be explained by other variables, such as gender. Gender differences have been found in relation to health behaviors in previous studies [84], and they have also been highlighted in work-life balance studies [5]. Some studies have found that women often have more difficulty maintaining a balance because of pressures at work and social role demands at home [85,86]. However, other studies have indicated that men and women report similar levels of work-life balance [87]. Other studies have shown that women tend to present a more positive work-life balance attitude compared with men [88]. Further studies that focus on the organization of work taking the employee's gender difference into account are recommended.

Moderation Role of Place of Residence

This study found that problematic internet use score in China was higher than that in Germany. This is in line with the previous studies, which found that people of an Asian background, such as Chinese, have higher problematic internet use scores than people from Europe or the United States [66,89].

Countries of residence rather than countries of birth were compared in this study, as it has been found that the effects on people's individual life from the place of residence appear to be higher than those from the place of birth [90]. From a closer look at country differences in health behaviors, it seems that the country where one lives provides the context within which a certain kind of behavior is shaped. In the case of Germany, it would be of interest to study the role of the country as a context variable in the health life and behavior of its residents, regardless of their cultural origin. Not only health behaviors but also work-life balance was different because of the country of residence, which involves influences of the natural environment, language, religion, custom, and particular ways of life [91,92]. This study found that the indirect association between health behaviors and work-life balance was moderated by the place of residence. The relation between health behaviors and work-life balance was significant in residents living in Germany, whereas it was weaker in residents in China.

This finding is in line with and further supports previous studies that explored health behaviors related to work-life balance [12,14,17,26,93]. This study also replicates studies that examined work-life balance across different societies [53]. The result showing that the relation between health behaviors and work-life balance was significant in Germany, whereas it was weaker in China, supports the *compensatory carry-over action model*. The model permits different applicability in 2 distinct countries of residence, from which, besides health behaviors, work-life balance could be affected by different factors. Previous studies have found that in Eastern countries, gender socialization plays an important role in one's perception toward work-life balance [53]. Moreover, although age was controlled in this study, other variables such as family responsibility status may be associated with work-life balance as well. Previous studies have indicated that older and married individuals felt less stressed than younger and unmarried people [75]. The possibility of married status in many Chinese participants might modulate the effect of health behaviors on work-life balance as well. This argument should be further explored in future research.

In this study, compared with residents in China, more people in Germany engaged in physical activity for at least 2.5 hours per week and drank enough water, whereas more people in Germany drank alcohol than people in China. These findings are in line with previous studies that have investigated health behaviors in Western countries and Eastern countries [63]. The moderation of the place of residence provides a broad perspective toward the promotion of work-life balance among 2 very diverse groups of countries of residence. By contrasting residents from China and Germany as representatives of prevailing Eastern and Western lifestyles, respectively, we may be able to gain a more global perspective in this issue. Professionals in the medical field should advise their patients in China to engage in more health behaviors and patients in Germany to find ways of lowering perceived stress to balance their lifestyle more adequately.

Limitations

Some limitations of this study should be addressed. First, this was a cross-sectional survey, which limits causal inferences. Moreover, there was a significant age difference between participants in the 2 groups. Although age and gender were included as control variables in this study, other variables such as family responsibility status may be associated with different ages and may impact the results as well. Second, health behaviors only included physical activity, nutrition and water intake, and smoking and alcohol status. However, other behaviors, such as dental hygiene adherence and sexual behaviors, are also related to health and may have an influence. Third, perceived stress was measured with only two items by using the shortened version of the measurement scale, and the Cronbach alpha was not high. The accuracy may not be as good as a full version. Finally, the data were mainly collected within Germany and China. Although participants with a migrant background were able to complete the English version of the questionnaire, the sample range was still limited. Moreover, the years of residence were not controlled, but it may also impact individuals' behaviors. Further studies should overcome these limitations by using longitudinal and experimental designs and considering more aspects and potential factors. It would also be valuable to include more people with different backgrounds regarding their country of residence while controlling other influences.

Conclusions

In summary, this study sheds light on how people's health behaviors interrelate with work-life balance, perceived stress, and problematic internet use. It provides evidence that health behaviors are indirectly related to work-life balance through perceived stress and problematic internet use. This study revealed a partial mediation among residents in Germany and a full mediation among residents in China. The relation between health behaviors and work-life balance was stronger in Germany,

whereas it was weaker in China. Besides performing health behaviors, helping people to relieve their perceived stress and spend only healthy amounts of time with the internet seems essential to maintain a better work-life balance. Future research should be performed in different countries and in different languages, and should be well grounded on a culture theory, such as Hofstede's cultural dimensions theory [48,49].

Acknowledgments

This research was partially funded by the Federal Ministry of Education and Research (BMBF) under research grant no. 01EL1822F and Jacobs University Bremen.

Conflicts of Interest

None declared.

References

1. Tanaka S, Maruyama Y, Ooshima S, Ito H. Working condition of nurses in Japan: awareness of work-life balance among nursing personnel at a university hospital. J Clin Nurs 2011 Jan;20(1-2):12-22. [doi: 10.1111/j.1365-2702.2010.03354.x] [Medline: 21083785]
2. Tuomi K, Vanhala S, Nykyri E, Janhonen M. Organizational practices, work demands and the well-being of employees: a follow-up study in the metal industry and retail trade. Occup Med (Lond) 2004 Mar;54(2):115-121. [doi: 10.1093/occmed/kqh005] [Medline: 15020730]
3. Albertsen K, Rafnsdóttir G, Grimsmo A, Tómasson K, Kauppinen K. Workhours and worklife balance. Scand J Work Environ Health 2008(5):14-21 [FREE Full text]
4. Jansen NW, Kant I, Nijhuis FJ, Swaen GM, Kristensen TS. Impact of worktime arrangements on work-home interference among Dutch employees. Scand J Work Environ Health 2004 Apr;30(2):139-148 [FREE Full text] [doi: 10.5271/sjweh.771] [Medline: 15143741]
5. Fagan C, Lyonette C, Smith M, Saldaña-Tejeda A, International Labour Office. Conditions of Work and Employment Branch. The Influence of Working Time Arrangements on Work-life Integration or 'Balance' : A Review of the International Evidence. Geneva: ILO; 2012.
6. Delecta P. Work life balance. Int J Curr Res 2011 Apr;3(4):186-189 [FREE Full text]
7. Yen JY, Yen CF, Chen CC, Chen SH, Ko CH. Family factors of internet addiction and substance use experience in Taiwanese adolescents. Cyberpsychol Behav 2007 Jun;10(3):323-329. [doi: 10.1089/cpb.2006.9948] [Medline: 17594255]
8. Ghassemzadeh L, Shahraray M, Moradi A. Prevalence of internet addiction and comparison of internet addicts and non-addicts in Iranian high schools. Cyberpsychol Behav 2008 Dec;11(6):731-733. [doi: 10.1089/cpb.2007.0243] [Medline: 18954277]
9. Brito AB, Bastos V, Junior SP, Franco J, Messias R, Pinho L, et al. Semantic Scholar. Montes Claros, Brazil: Unimontes Científica; 2018. Lifestyle and Internet Use: Addiction and Its Correlations URL: https://pdfs.semanticscholar.org/f995/a68a9d1e9d8a1f30e9a907ebaa7e488bd998.pdf?_ga=2.53645480.1035631690.1582702052-1679671381.1567599385 [accessed 2020-02-26]
10. Holden S, Sunindijo RY. Technology, long work hours, and stress worsen work-life balance in the construction industry. Int J Integr Eng 2018;10(2):-. [doi: 10.30880/ijie.2018.10.02.003]
11. Zschucke E, Renneberg B, Dimeo F, Wüstenberg T, Ströhle A. The stress-buffering effect of acute exercise: evidence for HPA axis negative feedback. Psychoneuroendocrinology 2015 Jan;51:414-425. [doi: 10.1016/j.psyneuen.2014.10.019] [Medline: 25462913]
12. Davis KD, Gere J, Sliwinski MJ. Investigating the work-family conflict and health link: repetitive thought as a mechanism. Stress Health 2017 Oct;33(4):330-338. [doi: 10.1002/smi.2711] [Medline: 27709792]
13. Fleig L, Küper C, Lippke S, Schwarzer R, Wiedemann AU. Cross-behavior associations and multiple health behavior change: a longitudinal study on physical activity and fruit and vegetable intake. J Health Psychol 2015 May;20(5):525-534. [doi: 10.1177/1359105315574951] [Medline: 25903240]
14. Mazzola JJ, Moore JT, Alexander K. Is work keeping us from acting healthy? How workplace barriers and facilitators impact nutrition and exercise behaviors. Stress Health 2017 Dec;33(5):479-489. [doi: 10.1002/smi.2731] [Medline: 27891758]
15. Rogers R, Prentice-Dunn S, Gochman D. Handbook of Health Behavior Research 1: Personal and Social Determinants. New York, NY, US: Plenum Press; 1997.
16. Lippke S, Nigg CR, Maddock JE. Health-promoting and health-risk behaviors: theory-driven analyses of multiple health behavior change in three international samples. Int J Behav Med 2012 Mar;19(1):1-13 [FREE Full text] [doi: 10.1007/s12529-010-9135-4] [Medline: 21234735]
17. Warnick JL, Pfammatter A, Champion K, Galluzzi T, Spring B. Perceptions of health behaviors and mobile health applications in an academically Elite College population to inform a targeted health promotion program. Int J Behav Med 2019 Apr;26(2):165-174. [doi: 10.1007/s12529-018-09767-y] [Medline: 30632092]

18. de Ridder D, Kroese F, Evers C, Adriaanse M, Gillebaart M. Healthy diet: health impact, prevalence, correlates, and interventions. Psychol Health 2017 Aug;32(8):907-941. [doi: 10.1080/08870446.2017.1316849] [Medline: 28447854]
19. Richert J, Reuter T, Wiedemann AU, Lippke S, Ziegelmann J, Schwarzer R. Differential effects of planning and self-efficacy on fruit and vegetable consumption. Appetite 2010 Jun;54(3):611-614. [doi: 10.1016/j.appet.2010.03.006] [Medline: 20227450]
20. de Graaf A, van den Putte B, Nguyen M, Zebregs S, Lammers J, Neijens P. The effectiveness of narrative versus informational smoking education on smoking beliefs, attitudes and intentions of low-educated adolescents. Psychol Health 2017 Jul;32(7):810-825. [doi: 10.1080/08870446.2017.1307371] [Medline: 28366018]
21. Chen X, Lau M, Kan MY, Chiang I, Hu Y, Gong J, et al. Lifestyle and addictive behaviors among Chinese adolescents in Hong Kong, Macau, Taipei, Wuhan, and Zhuhai - a first cross-subculture assessment. Int J Behav Med 2016 Oct;23(5):561-570. [doi: 10.1007/s12529-016-9548-9] [Medline: 26979424]
22. Black N, Mullan B, Sharpe L. Computer-delivered interventions for reducing alcohol consumption: meta-analysis and meta-regression using behaviour change techniques and theory. Health Psychol Rev 2016 Sep;10(3):341-357. [doi: 10.1080/17437199.2016.1168268] [Medline: 26999311]
23. Kuntsche E, Kuntsche S, Thrul J, Gmel G. Binge drinking: health impact, prevalence, correlates and interventions. Psychol Health 2017 Aug;32(8):976-1017. [doi: 10.1080/08870446.2017.1325889] [Medline: 28513195]
24. Keller S, Maddock JE, Hannöver W, Thyrian JR, Basler H. Multiple health risk behaviors in German first year university students. Prev Med 2008 Mar;46(3):189-195. [doi: 10.1016/j.ypmed.2007.09.008] [Medline: 18242666]
25. Boudreaux ED, Wood KB, Mehan D, Scarinci I, Taylor CL, Brantley PJ. Congruence of readiness to change, self-efficacy, and decisional balance for physical activity and dietary fat reduction. Am J Health Promot 2003;17(5):329-336. [doi: 10.4278/0890-1171-17.5.329] [Medline: 12769046]
26. Lippke S, Fleig L, Wiedemann AU, Schwarzer R. A computerized lifestyle application to promote multiple health behaviors at the workplace: testing its behavioral and psychological effects. J Med Internet Res 2015 Oct 1;17(10):e225 [FREE Full text] [doi: 10.2196/jmir.4486] [Medline: 26429115]
27. Lippke S. Modelling and supporting complex behavior change related to obesity and diabetes prevention and management with the compensatory carry-over action model. J Diabetes Obes 2014;1(2):1-5. [doi: 10.15436/2376-0494.14.009]
28. Koestner R, Lekes N, Powers TA, Chicoine E. Attaining personal goals: self-concordance plus implementation intentions equals success. J Pers Soc Psychol 2002 Jul;83(1):231-244. [Medline: 12088128]
29. Matuska K, Bass J, Schmitt JS. Life balance and perceived stress: predictors and demographic profile. OTJR (Thorofare N J) 2013;33(3):146-158. [doi: 10.3928/15394492-20130614-03] [Medline: 24651900]
30. Meyer D, Jayawardana MW, Muir SD, Ho DY, Sackett O. Promoting psychological well-being at work by reducing stress and improving sleep: mixed-methods analysis. J Med Internet Res 2018 Oct 19;20(10):e267 [FREE Full text] [doi: 10.2196/jmir.9058] [Medline: 30341045]
31. Malinauskas R. The associations among social support, stress, and life satisfaction as perceived by injured college athletes. Soc Behav Personal 2010;38(6):741-752. [doi: 10.2224/sbp.2010.38.6.741]
32. Warren CM, Otto AK, Walkner MM, Gupta RS. Quality of life among food allergic patients and their caregivers. Curr Allergy Asthma Rep 2016 May;16(5):38. [doi: 10.1007/s11882-016-0614-9] [Medline: 27048239]
33. Matuska K, Bass J. Life balance and stress in adults with medical conditions or obesity. OTJR (Thorofare N J) 2016 Apr;36(2):74-81. [doi: 10.1177/1539449216628859] [Medline: 27504880]
34. Sharma J, Dhar RL, Tyagi A. Stress as a mediator between work-family conflict and psychological health among the nursing staff: Moderating role of emotional intelligence. Appl Nurs Res 2016 May;30:268-275. [doi: 10.1016/j.apnr.2015.01.010] [Medline: 25769936]
35. Allen TD, Armstrong J. Further examination of the link between work-family conflict and physical health. Am Behav Sci 2006;49(9):1204-1221. [doi: 10.1177/0002764206286386]
36. Lippke S, Wienert J, Kuhlmann T, Fink S, Hambrecht R. Perceived stress, physical activity and motivation: findings from an internet study. Ann Sports Med Re 2015;2(1):1012-1017 [FREE Full text]
37. Milkie MA, Mattingly MJ, Nomaguchi KM, Bianchi SM, Robinson JP. The time squeeze: parental statuses and feelings about time with children. J Marriage Fam 2004 Aug;66(3):739-761. [doi: 10.1111/j.0022-2445.2004.00050.x]
38. Widyanto L, Griffiths M. Unravelling the web: adolescents and internet addiction. In: Information Resources Management Association , editor. Virtual Communities: Concepts, Methodologies, Tools and Applications. Hershey, Pennsylvania, USA: IGI Global; 2011:2433-2453.
39. Ioannidis K, Treder MS, Chamberlain SR, Kiraly F, Redden SA, Stein DJ, et al. Problematic internet use as an age-related multifaceted problem: evidence from a two-site survey. Addict Behav 2018 Jun;81:157-166 [FREE Full text] [doi: 10.1016/j.addbeh.2018.02.017] [Medline: 29459201]
40. Akin A, Iskender M. Internet addiction and depression, anxiety and stress. Int Online J Educ Sci 2011;3(1):138-148 [FREE Full text]
41. Samaha M, Hawi NS. Relationships among smartphone addiction, stress, academic performance, and satisfaction with life. Comput Human Behav 2016;57:321-325. [doi: 10.1016/j.chb.2015.12.045]

42. Ostovar S, Allahyar N, Aminpoor H, Moafian F, Nor MB, Griffiths MD. Internet addiction and its psychosocial risks (depression, anxiety, stress and loneliness) among Iranian adolescents and young adults: A structural equation model in a cross-sectional study. Int J Ment Health Addiction 2016;14(3):257-267. [doi: 10.1007/s11469-015-9628-0]
43. Khalili-Mahani N, Smyrnova A, Kakinami L. To each stress its own screen: a cross-sectional survey of the patterns of stress and various screen uses in relation to self-admitted screen addiction. J Med Internet Res 2019 Apr 2;21(4):e11485 [FREE Full text] [doi: 10.2196/11485] [Medline: 30938685]
44. Li D, Zhang W, Li X, Zhou Y, Zhao L, Wang Y. Stressful life events and adolescent Internet addiction: the mediating role of psychological needs satisfaction and the moderating role of coping style. Comput Human Behav 2016;63(17):408-415. [doi: 10.1016/j.chb.2016.05.070]
45. Zhao F, Zhang Z, Bi L, Wu X, Wang W, Li Y, et al. The association between life events and internet addiction among Chinese vocational school students: the mediating role of depression. Comput Human Behav 2017;70:30-38. [doi: 10.1016/j.chb.2016.12.057]
46. Schwartz SH. Individualism-collectivism: critique and proposed refinements. J Cross Cult Psychol 1990;21(2):139-157. [doi: 10.1177/0022022190212001]
47. Oetzel J, Ting-Toomey S, Masumoto T, Yokochi Y, Pan X, Takai J, et al. Face and facework in conflict: a cross-cultural comparison of China, Germany, Japan, and the United States. Commun Monogr 2001;68(3):235-258. [doi: 10.1080/03637750128061]
48. Hofstede G. Culture and organizations. Int Stud Manag Organ 1980;10(4):15-41. [doi: 10.1080/00208825.1980.11656300]
49. Hofstede G. Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations. Thousand Oaks, California, USA: Sage publications; 2001.
50. Diehl S, Terlutter R, Weinberg P. Advertising Effectiveness in Different Cultures: Results of an Experiment Analyzing the Effects of Individualistic and Collectivistic Advertising on Germans and Chinese. In: Proceedings of the 2003 European Advances in Consumer Research. 2003 Presented at: ACR'03; June 4-7, 2003; Dublin.
51. Singelis TM, Triandis HC, Bhawuk DP, Gelfand MJ. Horizontal and vertical dimensions of individualism and collectivism: a theoretical and measurement refinement. Cross-Cult Res 1995;29(3):240-275. [doi: 10.1177/106939719502900302]
52. Earley PC. Social loafing and collectivism: a comparison of the United States and the People's Republic of China. Adm Sci Q 1989;34(4):565-581. [doi: 10.2307/2393567]
53. Chandra V. Work–life balance: eastern and western perspectives. Int J Hum Res Manag 2012 Mar;23(5):1040-1056. [doi: 10.1080/09585192.2012.651339]
54. Xiao Y, Cooke FL. Work-life balance in China? Social policy, employer strategy and individual coping mechanisms. Asia Pac J Hum Res 2012;50(1):6-22. [doi: 10.1111/j.1744-7941.2011.00005.x]
55. Lu J, Siu O, Spector PE, Shi K. Antecedents and outcomes of a fourfold taxonomy of work-family balance in Chinese employed parents. J Occup Health Psychol 2009 Apr;14(2):182-192. [doi: 10.1037/a0014115] [Medline: 19331479]
56. Gan Y, Gan T, Chen Z, Miao M, Zhang K. Functional pathways of social support for mental health in work and family domains among Chinese scientific and technological professionals. Stress Health 2015 Oct;31(4):336-349. [doi: 10.1002/smi.2572] [Medline: 24782044]
57. Lu L, Cooper CL, Kao S, Chang T, Allen TD, Lapierre LM, et al. Cross-cultural differences on work-to-family conflict and role satisfaction: a Taiwanese-British comparison. Hum Res Manage 2010;49(1):67-85. [doi: 10.1002/hrm.20334]
58. Sorensen G, Emmons K, Hunt MK, Barbeau E, Goldman R, Peterson K, et al. Model for incorporating social context in health behavior interventions: applications for cancer prevention for working-class, multiethnic populations. Prev Med 2003 Sep;37(3):188-197. [doi: 10.1016/s0091-7435(03)00111-7] [Medline: 12914824]
59. Sansone G, Fong GT, Hall PA, Guignard R, Beck F, Mons U, et al. Time perspective as a predictor of smoking status: findings from the International Tobacco Control (ITC) Surveys in Scotland, France, Germany, China, and Malaysia. BMC Public Health 2013 Apr 15;13:346 [FREE Full text] [doi: 10.1186/1471-2458-13-346] [Medline: 23587205]
60. Siegrist M, Shi J, Giusto A, Hartmann C. Worlds apart. Consumer acceptance of functional foods and beverages in Germany and China. Appetite 2015 Sep;92:87-93. [doi: 10.1016/j.appet.2015.05.017] [Medline: 26002279]
61. Wright LT, Nancarrow C, Kwok PM. Food taste preferences and cultural influences on consumption. Br Food J 2001;103(5):348-357. [doi: 10.1108/00070700110396321]
62. Hulteen RM, Smith JJ, Morgan PJ, Barnett LM, Hallal PC, Colyvas K, et al. Global participation in sport and leisure-time physical activities: a systematic review and meta-analysis. Prev Med 2017 Feb;95:14-25. [doi: 10.1016/j.ypmed.2016.11.027] [Medline: 27939265]
63. Grant N, Wardle J, Steptoe A. The relationship between life satisfaction and health behavior: a cross-cultural analysis of young adults. Int J Behav Med 2009;16(3):259-268. [doi: 10.1007/s12529-009-9032-x] [Medline: 19319695]
64. Chun CA, Moos RH, Cronkite RC. Culture: a fundamental context for the stress and coping paradigm. In: Wong PT, Wong LC, editors. Handbook of Multicultural Perspectives on Stress and Coping. Salmon Tower Building, New York City, USA: Springer; 2006:29-53.
65. Sawang S, Oei TP, Goh YW. Are Country and Culture Values Interchangeable? Int J Cross Cult Manag 2006;6(2):205-219. [doi: 10.1177/1470595806066330]

66. Lopez-Fernandez O. Cross-cultural research on internet addiction: a systematic review. Int Arch Addict Res Med 2015;1(2):1-8. [doi: 10.23937/2474-3631/1510011]
67. Peterka-Bonetta J, Sindermann C, Sha P, Zhou M, Montag C. The relationship between Internet Use Disorder, depression and burnout among Chinese and German college students. Addict Behav 2019 Feb;89:188-199. [doi: 10.1016/j.addbeh.2018.08.011] [Medline: 30321691]
68. Webb A. Bloomberg. 2014 May 20. Germany Top Migration Land After US in New OECD Ranking URL: https://www.bloomberg.com/news/articles/2014-05-20/immigration-boom-propels-germany-past-u-k-in-new-oecd-ranking [accessed 2019-09-01]
69. Mäder U, Martin BW, Schutz Y, Marti B. Validity of four short physical activity questionnaires in middle-aged persons. Med Sci Sports Exerc 2006 Jul;38(7):1255-1266. [doi: 10.1249/01.mss.0000227310.18902.28] [Medline: 16826022]
70. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
71. Lippke S, Ziegelmann JP, Schwarzer R, Velicer WF. Validity of stage assessment in the adoption and maintenance of physical activity and fruit and vegetable consumption. Health Psychol 2009 Mar;28(2):183-193 [FREE Full text] [doi: 10.1037/a0012983] [Medline: 19290710]
72. Renner B, Hahn A, Schwarzer R. Risiko und Gesundheitsverhalten. Berlin, Germany: Free University Berlin; 1996.
73. Syrek C, Bauer-Emmel C, Antoni C, Klusemann J. Entwicklung und Validierung der Trierer Kurzskala zur Messung von Work-Life Balance (TKS-WLB). Diagnostica 2011 Jul;57(3):134-145. [doi: 10.1026/0012-1924/a000044]
74. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [doi: 10.2307/2136404] [Medline: 6668417]
75. Klein EM, Brähler E, Dreier M, Reinecke L, Müller KW, Schmutzer G, et al. The German version of the Perceived Stress Scale - psychometric characteristics in a representative German community sample. BMC Psychiatry 2016 May 23;16:159 [FREE Full text] [doi: 10.1186/s12888-016-0875-9] [Medline: 27216151]
76. Hahn A, Jerusalem M, Meixner-Dahle S. André Hahn. 2014. Fragebogen zur Internetsucht (ISS-20r) und Normwerte URL: http://andre-hahn.de/downloads/pub/2014/2014_ISS20-Internetsuchtskala.pdf [accessed 2019-09-10]
77. Hayes AF. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach. New York City, USA: Guilford Publications; 2017.
78. Hayes AF. Andrew F Hayes. 2012. PROCESS: A Versatile Computational Tool for Observed Variable Mediation, Moderation, and Conditional Process Modeling White Paper URL: http://afhayes.com/public/process2012.pdf [accessed 2019-09-10]
79. English AS, Zeng ZJ, Ma JH. The stress of studying in China: primary and secondary coping interaction effects. Springerplus 2015;4:755 [FREE Full text] [doi: 10.1186/s40064-015-1540-3] [Medline: 26693113]
80. Quiñones-García C, Korak-Kakabadse N. Compulsive internet use in adults: a study of prevalence and drivers within the current economic climate in the UK. Comput Human Behav 2014;30:171-180. [doi: 10.1016/j.chb.2013.08.004]
81. Buckner JE, Castille CM, Sheets TL. The Five Factor Model of personality and employees' excessive use of technology. Computers in Human Behavior 2012;28(5):1947-1953. [doi: 10.1016/j.chb.2012.05.014]
82. Goodman CA. Digital Commons @ Gardner-Webb University. Digital Commons @ Gardner-Webb University: Gardner-Webb; 2017. Stress and Social Media: Implications for Nurses' Quality of Life URL: https://digitalcommons.gardner-webb.edu/cgi/viewcontent.cgi?article=1272&context=nursing_etd [accessed 2019-09-10]
83. Sturgeon JA, Arewasikporn A, Okun MA, Davis MC, Ong AD, Zautra AJ. The psychosocial context of financial stress: Implications for inflammation and psychological health. Psychosom Med 2016;78(2):134-143 [FREE Full text] [doi: 10.1097/PSY.0000000000000276] [Medline: 26569541]
84. Minges KE, Chao A, Nam S, Grey M, Whittemore R. Weight status, gender, and race/ethnicity: are there differences in meeting recommended health behavior guidelines for adolescents? J Sch Nurs 2015 Apr;31(2):135-145 [FREE Full text] [doi: 10.1177/1059840514554089] [Medline: 25312400]
85. Rehman S, Roomi MA. Gender and work‐life balance: a phenomenological study of women entrepreneurs in Pakistan. J Small Bus Enterp Dev 2012;19(2):209-228. [doi: 10.1108/14626001211223865]
86. Griep RH, Toivanen S, van Diepen C, Guimarães JM, Camelo LV, Juvanhol LL, et al. Work-family conflict and self-rated health: the role of gender and educational level. Baseline data from the Brazilian longitudinal study of adult health (ELSA-Brasil). Int J Behav Med 2016 Jun;23(3):372-382 [FREE Full text] [doi: 10.1007/s12529-015-9523-x] [Medline: 26597924]
87. Emslie C, Hunt K, Macintyre S. Gender, work-home conflict, and morbidity amongst white-collar bank employees in the United Kingdom. Int J Behav Med 2004;11(3):127-134. [doi: 10.1207/s15327558ijbm1103_1] [Medline: 15496340]
88. Selvarajan T, Slattery J, Stringer DY. Relationship between gender and work related attitudes: a study of temporary agency employees. J Bus Res 2015;68(9):1919-1927. [doi: 10.1016/j.jbusres.2015.01.001]
89. Samli R. A review of internet addiction on the basis of different countries (2007–2017). In: Information Resources Management Association , editor. Substance Abuse and Addiction: Breakthroughs in Research and Practice. Commonwealth of Pennsylvania, USA: IGI Global; 2019:215-235.

90. Rudnev M. Value adaptation among intra-European migrants. J Cross Cult Psychol 2014 Aug;45(10):1626-1642. [doi: 10.1177/0022022114548482]
91. Benson M, O'Reilly K. Migration and the search for a better way of life: a critical exploration of lifestyle migration. Soc Rev 2009;57(4):608-625. [doi: 10.1111/j.1467-954x.2009.01864.x]
92. Hossain K. Securitizing the Arctic indigenous peoples: a community security perspective with special reference to the Sámi of the European high north. Polar Sci 2016;10(3):415-424. [doi: 10.1016/j.polar.2016.04.010]
93. van Scheppingen AR, de Vroome EM, ten Have KC, Zwetsloot GI, Bos EH, van Mechelen W. Motivations for health and their associations with lifestyle, work style, health, vitality, and employee productivity. J Occup Environ Med 2014 May;56(5):540-546. [doi: 10.1097/JOM.0000000000000143] [Medline: 24806567]

Abbreviations

ISS-10r: Internet Addiction Questionnaire

Edited by G Eysenbach; submitted 02.10.19; peer-reviewed by N Khalili-Mahani, M Wong, Z Ma; comments to author 27.11.19; revised version received 09.12.19; accepted 07.02.20; published 11.05.20.

Please cite as:
Gao L, Gan Y, Whittal A, Yan S, Lippke S
The Mediator Roles of Problematic Internet Use and Perceived Stress Between Health Behaviors and Work-Life Balance Among Internet Users in Germany and China: Web-Based Cross-Sectional Study
J Med Internet Res 2020;22(5):e16468
URL: https://www.jmir.org/2020/5/e16468
doi:10.2196/16468
PMID:32391798

Original Paper

A Cross-Cultural Comparison of an Extended Planned Risk Information Seeking Model on Mental Health Among College Students: Cross-Sectional Study

Zhaomeng Niu[1], PhD; Jessica Fitts Willoughby[2], PhD; Jing Mei[3], PhD; Shaochun Li[3], MSc; Pengwei Hu[3], PhD

[1]Rutgers Cancer Institute of New Jersey, New Brunswick, NJ, United States

[2]Washington State University, Pullman, WA, United States

[3]AI for Healthcare, IBM Research, Beijing, China

Corresponding Author:
Pengwei Hu, PhD
AI for Healthcare
IBM Research
19 Zhonguancun Software Park
8 Dong Bei Wang Xi Lu, Haidian Qu,
Beijing, 100085
China
Phone: 86 10 58748625
Email: hupwei@cn.ibm.com

Abstract

Background: Approximately 42.5 million adults have been affected by mental illness in the United States in 2013, and 173 million people have been affected by a diagnosable psychiatric disorder in China. An increasing number of people tend to seek health information on the Web, and it is important to understand the factors associated with individuals' mental health information seeking. Identifying factors associated with mental health information seeking may influence the disease progression of potential patients. The planned risk information seeking model (PRISM) was developed in 2010 by integrating multiple information seeking models including the theory of planned behavior. Few studies have replicated PRISM outside the United States and no previous study has examined mental health as a personal risk in different cultures.

Objective: This study aimed to test the planned risk information seeking model (PRISM) in China and the United States with a chronic disease, mental illness, and two additional factors, ie, media use and cultural identity, among college students.

Methods: Data were collected in both countries using the same online survey through a survey management program (Qualtrics). In China, college instructors distributed the survey link among university students, and it was also posted on a leading social media site called Sina Weibo. In the United States, the data were collected in a college-wide survey pool in a large Northwestern university.

Results: The final sample size was 235 for the Chinese sample and 241 for the US sample. Media use was significantly associated with mental health information–seeking intentions in the Chinese sample (P<.001), and cultural identity was significantly associated with intentions in both samples (China: P=.02; United States: P<.001). The extended PRISM had a better model fit than the original PRISM.

Conclusions: Cultural identity and media use should be considered when evaluating the process of mental health information seeking or when designing interventions to address mental health information seeking.

(J Med Internet Res 2020;22(5):e15817) doi:10.2196/15817

KEYWORDS

information seeking behavior; mental health; cross-cultural comparison

Introduction

Background

In the United States, about 42.5 million adults have been affected by various kinds of mental illness in 2013, which means that 1 in 5 Americans have had a diagnosable mental disorder [1]. In China, approximately 173 million people experience a diagnosable psychiatric disorder [2]. People in the United States who are not mental health professionals tend to recognize serious mental issues, such as schizophrenia, as mental illness. Mental disorders, such as depression and paranoia, which also require proper treatment, are often not perceived as mental disorders [3,4]. The situation is worse in China; the majority of people have a negative attitude toward mental health counseling services [5]. Some Asians see admitting to having a mental illness as bringing disgrace to the family and believe that willpower could maintain mental health to some extent [6,7]. Traditional Chinese culture and the history of medicine do not define mental illness separately from neurological disorders [8], whereas in the United Sates, psychiatrists are trying to incorporate medical treatment with counseling services that can help analyze the social and environmental factors of patients.

Moreover, Chinese culture has been known as a collectivistic culture, where group members traditionally put group interests and community values ahead of individual interests [9]. Therefore, people who suffer from mental illness may refuse to seek counseling services because of the concern that they may bring shame to the family or community. In the United States as well, there is stigma around mental health. Even among professional health care workers, negative attitudes toward mental illness are pervasive [10]. Therefore, individuals may avoid interpersonal communication and turn to the internet for information related to mental health.

With the ubiquitous use of the internet (eg, used by more than 3 billion people throughout the world) [11], people now have unparalleled access to information. In the United States, 72.00% of internet users have sought health information online in 2013[12]. In China, research has found that more than 80% of individuals experiencing a specific health condition (ie, people with epilepsy) have sought information related to their illness [13]. One study found that 18% of adult internet users in the United States had searched for mental health information online [14]. Using the internet for mental health information has been found to be prevalent among young adults. Horgan and Sweeney [15] found that 30.8% of their participants, who were college students, had used the internet for information related to mental health, conducting searches mainly on depression. This may be tied to the age of onset for mental illness; three-quarters of all chronic mental illnesses begins by the age of 24 years [16]. College may also be a risky time for students' mental health, and college students tend to feel depressed and anxious during their college life [17]. Previous studies have found that more than 20% of college students reported mental health problems in both China [18] and the United State [19], and they are also heavy media users [20]. Thus, mental health information seeking is of importance among a college population. Therefore, a college student sample was used for this study. As the United States and China are two typical but distinct cultures that can represent different populations in mental health information seeking, a cross-cultural study was conducted in these 2 countries among college students.

The Planned Risk Information Seeking Model

There has been a fair amount of studies about health-related information seeking since the 1990s [21]. In 2010, Kahlor [22] developed the planned risk information seeking model (PRISM) by integrating multiple information seeking models including the theory of planned behavior [23], the risk information seeking and processing (RISP) model [24], and the augmented RISP model [25]. PRISM [22] proposes that subjective norms, attitude toward seeking, and perceived seeking control will be positively associated with seeking intent; affective response to risk and perceived knowledge insufficiency will be positively related with seeking intent; perceived knowledge partially mediates the effects of attitude, seeking-related subjective norms, and perceived control on perceived knowledge insufficiency; and risk perception will be positively related to affective risk response.

When proposing PRISM, Kahlor [22] assessed the topic of general health risk information seeking, finding the model a good fit for the data, although the predicted path of knowledge insufficiency to seeking intent was not supported. Additional research has found support for PRISM to be an acceptable model for predicting information seeking intentions [26-28]. Hovick et al [29] found that original variables in PRISM better fit their data than an expanded model with past seeking behavior, source beliefs, and outcome expectancies regarding the context of cancer risk. However, Ho et al [27] extended PRISM in Singapore by adding media use and found the extended PRISM had a better fit than the PRISM. Few studies have replicated PRISM outside the United States and/or compared the relationships within PRISM under different cultural contexts. Furthermore, Willoughby and Myrick [28] found, in an examination of PRISM in two health contexts, that although some of the paths were not supported in the model, PRISM was a good fit for two health contexts: sexual health and cancer. No previous study has examined mental health as a personal risk in different cultures. This study aimed to examine whether PRISM is a good fit for the data in both China and the United States regarding mental health.

Media Use

Although PRISM incorporates a number of constructs from various theories, it does not include the construct of media use. Media use can serve as a source of knowledge and provide important information; it can also be affected by knowledge needs and remind information seekers to stay alert. Previous research found media use to be a mediator in PRISM for predicting impersonal risk information seeking, specifically examining climate change [27]. Media use has not been assessed as a predictor in PRISM when examining personal risks. Personal risks may be more emotional than impersonal risks [30]. This study used a personal risk, mental health, as a context to replicate and extend the current model.

Demographics, ideology, personality traits, and social persuasive effects influence media use [31]. According to the reinforcing spirals framework [31], media can influence attitudes and behaviors; furthermore, media use can also influence future media use or avoidance. In information seeking, media are often seen as information sources. Media use has been found to be associated with health information seeking for impersonal risk [27,32]. In addition, people who engage with one media channel for health information are more likely to engage with other media channels in their search for health information [33].

The reinforcing spirals framework highlights the potential timeline and mechanisms for how media use might work. For example, a person who has had a family member experience depression might be primed to have certain attitudes and opinions. After seeing a commercial for a medication for depression, the person might be more likely to seek out additional information on signs and symptoms because of their own personal past experiences and also the information they were exposed to through the media.

A number of studies have shown that media use could influence perceived knowledge and intentions within the context of health [34-36]. Little research has examined the relationship between media use and variables within PRISM under the context of mental health.

Cultural Identity

Another factor not included in PRISM is cultural identity, which is identification with a particular cultural group [37]. Even with the same cultural background, people can have different levels of cultural identity. Few studies about information seeking have examined the role of cultural values and how different levels of cultural identity affect the process of information seeking. One cross-cultural study discovered that people from different cultures, such as a communitarian culture (eg, Chinese culture) and an individualistic culture (eg, US culture), tend to perceive variables differently in risk information seeking [38]. Although this research assessed risk information seeking across cultures, it did not include cultural identify as a predictor. Our study assesses the potential role of cultural identity as a predictor of mental health information seeking.

Hofstede [39] defines culture as "the collective programming of the mind, which distinguishes one group or category of people from another." Even under the same cultural context, individuals can experience different levels of cultural identity. Cultural identity was defined as an individual's identification with a particular cultural group [37]. A strong cultural identity is accompanied with a deep understanding of group values, social norms, and certain behaviors endorsed by the community members [37]. For example, Chinese culture values tend to stigmatize mental health and not talk about it. If someone instead believed it was normal to have mental issues and believed it was healthy to talk with a therapist, this person's cultural identity of Chinese culture regarding mental health would be low.

Chinese culture, as a typical Asian culture, is known as highly collectivistic with value placed on group members' opinions and a focus on cooperation in group settings [39,40]. There is evidence that Chinese value social norms more and are more likely to maintain the existing social structure than people from individualistic cultures [41,42]. Therefore, we believe that the seeking intent of mental health information will be weaker among a Chinese sample not only because of the previous treatment of mental health in China but also because many individuals from a collectivistic culture see having mental health issues as losing face or bringing damage to the family's reputation [6]. Moreover, the cultural value of the United States is recognized as individualistic [39], and people from individualistic cultures are more likely to accept mental health counseling [6].

Research Questions and Hypotheses

On the basis of the previously discussed literature, we proposed the following research questions and hypotheses to help assess how the additional variables may impact behavioral intentions and the utility of the PRISM model at predicting behavioral intentions in different samples:

- Research question (RQ) 1: Is PRISM a good fit for both the Chinese sample and the US sample of young adults on the topic of mental health information seeking?
- Hypothesis 1: Media use will be positively related to seeking intention in the Chinese (H1a) and US (H1b) samples.
- Hypothesis 2: Cultural identity will be negatively associated with mental health information seeking intentions among participants in the Chinese sample.
- Hypothesis 3: Cultural identity will be positively associated with mental health information seeking intentions among participants in the US sample.
- RQ2: What are the relationships between media use and perceived knowledge, perceived knowledge insufficiency, affective response, risk perception, attitude, and subjective norms in the PRISM in the mental health information seeking process in both samples?
- RQ3: What are the relationships between cultural identity and variables of PRISM in the mental health information seeking process in both samples?
- RQ4: Does the extended PRISM account for more variance in seeking intentions than the PRISM in both the Chinese (RQ4a) and US (RQ4b) samples?

Methods

Sample and Procedure

Data were collected in both countries using the same online survey through a survey management program (Qualtrics). The studies were similar in all aspects, except during the recruitment phase. Participants in the United States were recruited from a college participation system called Sona, a cloud-based subject pool. Participants in China were recruited from 2 universities in China and a social media site. Although the methods differ, both provided a sample of college students that was obtained as a convenience sample. These differences in recruitment were due, in part, to logistical issues. The researchers did not have access to a participant pool in China but wanted a comparable population of college students in the same age range. The University Institutional Review Board reviewed the study

proposal, and the project was determined to be exempt for both samples.

In China, 2 college instructors in 2 different universities distributed the anonymous survey link to their students using online class announcements and the survey link was also posted on a leading social media site called Sina Weibo to ensure an equivalent sample size. Sina Weibo is often seen as the "Chinese Twitter" and is a microblogging website in China. The questionnaire was translated into Chinese and was cross-examined by an English instructor of a university in China to verify whether the 2 survey versions were consistent. Two students volunteered to pretest the survey, and the completion time was around 15 min. Respondents needed to be 18 years or older to participate. One researcher posted the study description including study information, target population, survey link, and incentive information on Sina Weibo, and snowball sampling was used to distribute the survey link. At the beginning of the online survey, students needed to give their consent to participate in the study by selecting "I am 18 or older and agree to participate." An incentive of CNY ¥260 (US $37.02) or an equivalent prize was provided in the form of a random drawing. At the end of the survey, participants were provided with another link to enter their email address for a chance to win the incentive. The email addresses were only used for the draw, and participants could not be identified by either the researchers or the instructors. The data collection lasted for about a month, and the completion rate was 77.6%. Incomplete questionnaires and surveys that took less than 2 min to complete were not included in the final analysis.

In the United States, the data were collected in a college-wide subject pool (Sona) in a large Northwestern university. Respondents were recruited among undergraduate students from different majors, and participants were offered extra credit for participating in the study. Participants signed up in Sona and then were directed to the survey in Qualtrics. Participants needed to give consent before they could proceed to the main questionnaire. The data collection process lasted less than 2 months. The completion rate was 96.78%. Students had options for alternate assignments to receive similar credit for their courses. Questionnaires that were incomplete and took less than 2 min to finish were not included in the final analysis.

Measurement

We used the existing measures of attitude toward seeking, subjective norms, perceived seeking control, perceived current knowledge, risk perception, affective response, perceived knowledge insufficiency, and seeking intent from Kahlor [22].

Attitude Toward Seeking

Seven 7-point scale statements were used to measure respondents' attitude toward seeking mental health information. Questions asked whether seeking mental health–related information was "bad" or "good," "unhelpful" or "helpful," "worthless" or "valuable," "unproductive" or "productive," "harmful" or "beneficial," "foolish" or "wise," and "not useful" or "useful." Items were averaged to create a scale (United States: alpha=.95, mean 5.91, SD 1.11; China: alpha=.90, mean 5.56, SD 0.97).

Seeking-Related Subjective Norms

Five 5-point Likert-type items measured the degree of agreement with statements regarding subjective norms (eg, "Most people who are important to me think that I should seek information about risks to my mental health.") Items were averaged to create a scale (United States: alpha=.92, mean 2.81, SD 1.13; China: alpha=.91, mean 2.62, SD 1.03).

Perceived Seeking Control

Four 5-point Likert-type items measured the degree of agreement with statements regarding perceived seeking control (eg, "I can readily access all the information about risks to my mental health that I need"). Items were averaged to create a scale (United States: alpha=.90, mean 3.47, SD 0.91; China: alpha=.91, mean 3.23, SD 0.98).

Risk Perception

Three 11-point items measured risk perception related to mental health (0=not at all and 10=extremely) with the following statements: "How serious are the current threats to your mental health?" "How likely are you to have some mental health issues in the next year?" and "If you were to have some mental health issues in the next year, how serious do you think it would be?" Items were averaged to create a scale (United States: alpha=.88, mean 3.86, SD 2.31; China: alpha=.88, mean 5.32, SD 2.30).

Affective Response

Two 5-point Likert-type items asked respondents to indicate their degree of worry and fear. The statements were "Current risks to my mental health are scary" and "Current risks to my mental health are worrisome." The items were averaged to create a scale (United States: alpha=.94, mean 2.39, SD 1.26; China: alpha=.89, mean 2.76, SD 1.14).

Perceived Current Knowledge

A statement measured perceived current knowledge by asking respondents the following: "Rate your mental health risk knowledge on a scale of 0 to 100, where zero means knowing nothing about risks to your mental health and 100 means knowing everything you could possibly know about risks to your mental health" (United States: mean 57.48, SD 24.30; China: mean 64.09, SD 18.06).

Perceived Knowledge Insufficiency

The measurement of sufficiency threshold asked the following: "Think of that same 0 to 100 scale again. This time, estimate how much knowledge you need to deal adequately with risks to your mental health. You might feel you need the same, more, or possibly even less information about this topic. Using a scale of zero to 100, how much information would be sufficient for you" (United States: mean 66.14, SD 23.45; China: mean 74.92, SD 17.34).

Seeking Intentions

Five 5-point Likert-type items measured seeking intent (eg, "I plan to seek more information about risks to my mental health in the near future"). Items were averaged to create a scale (United States: alpha=.97, mean 2.90, SD 1.07; China: alpha=.94, mean 2.98, SD 1.01).

Media Use

We adapted Brossard and Nisbet's [43] scale. Eight 7-point items measured participants' attention paid to mental health information in health magazines, newspapers, TV news reports, entertainment TV programs, online forums, social media, news reports on the internet, and medical applications (eg, "How much attention have you paid to mental health information on social media?"). Items were averaged to create a scale (United States: alpha=.82, mean 3.34, SD 1.17; China: alpha=.84, mean 3.47, SD 1.22).

Cultural Identity

We measured cultural identity with Usborne and Taylor's [37] Cultural Identity Clarity Scale. Eight 11-point items measured respondents' agreement or disagreement with statements such as "My beliefs about my cultural group often conflict with one another" and "My beliefs about my cultural group seem to change very frequently." Some items were reverse coded, and all items were averaged to create a scale (United States: alpha=.87, mean 5.68, SD 1.89; China: alpha=.76, mean 5.33, SD 1.67).

Data Analysis

As we wanted to examine a college student sample across both countries, we only used those who were college students in the both samples. As the measurements were used in former studies and have high reliability (high Cronbach alpha), we computed the items to create one variable for each construct and then conducted path analyses in Mplus version 7.11 to evaluate the paths and model fit. The models included gender as a control variable.

Results

Sample Statistics

After data cleaning, each sample had more than 200 participants complete the survey (United States: N=241; China: N=235). In the US sample, more than one-third of the sample reported being male (n=83) and less than three-quarters, female (n=158). Participants' ages ranged from 18 to 32 years (mean 20 years, SD 1.97). Almost two-thirds of the respondents (n=160) reported being white and >10.0% (n=27) as Asian. In the Chinese sample, almost 90.0% (n=209) were female. Respondents ranged from 18 to 27 years (mean 21 years, SD 7.98). Basic descriptive results of the variables are shown in Table 1.

Table 1. Basic descriptive results.

Variable	United States, mean (SD)	China, mean (SD)	Range
Attitude toward seeking	5.91 (1.11)	5.56 (0.97)	1-7
Seeking-related subjective norms	2.81 (1.13)	2.62 (1.03)	1-5
Perceived seeking control	3.47 (0.91)	3.23 (0.98)	1-5
Risk perception	3.86 (2.31)	5.32 (2.30)	0-10
Affective response	2.39 (1.26)	2.76 (1.14)	1-5
Perceived current knowledge	57.48 (24.30)	64.09 (18.06)	0-100
Perceived knowledge insufficiency	66.14 (23.45)	74.92 (17.34)	0-100
Seeking intentions	2.90 (1.07)	2.98 (1.01)	1-5
Media use	3.34 (1.17)	3.47 (1.22)	1-7
Cultural identity	5.68 (1.89)	5.33 (1.67)	0-10

Model Fit

For the Chinese sample, the replicated PRISM did not have a good fit, whereas the extended model with 2 additional variables, attention to media and cultural identity, had a good model fit. The replicated PRISM of the US participants did not have a good model fit, whereas the extended PRISM of the US sample had an acceptable model fit (see Table 2 for all model fit information). According to Browne and Cudeck [44], root mean squared error of approximation values less than 0.08 can be considered as the reasonable error of approximation. A comparative fit index and Tucker-Lewis index close to or greater than 0.95 represent a good fit. A normed chi-square value less than 5 indicates a good fit [45]. Thus, RQ1 was answered.

Table 2. Summary of model fit.

Model	Chi-square (*df*)	Root mean squared error of approximation	Comparative Fit Index	Tucker-Lewis Index
China PRISM[a]	55.4 (2.41)	0.078	0.91	0.89
China Extended PRISM	57.1 (1.59)	0.050	0.96	0.93
US PRISM	90.1 (3.91)	0.110	0.77	0.70
US Extended PRISM	75.4 (2.09)	0.067	0.92	0.86

[a]PRISM: Planned Risk Information Seeking Model.

The Planned Risk Information Seeking Model of the Chinese Sample

Some significant paths of Kahlor's [22] PRISM were not significant in the Chinese sample (Figure 1) Perceived seeking control was not significantly related to seeking intention, seeking-related subjective norms were not significantly related to perceived information insufficiency, and seeking-related subjective norms were not significantly related to perceived current knowledge.

The Planned Risk Information Seeking Model of the US Sample

The paths and standardized coefficients of the replicated PRISM of the US sample are presented in Figure 2.

Figure 1. The planned risk information seeking model of the Chinese sample. Dashed lines denote hypothesized nonsignificant paths. The model includes effects of control variables, which are not displayed. (.000) represents significant path coefficients at the .001 level and (.01), at the .05 level.

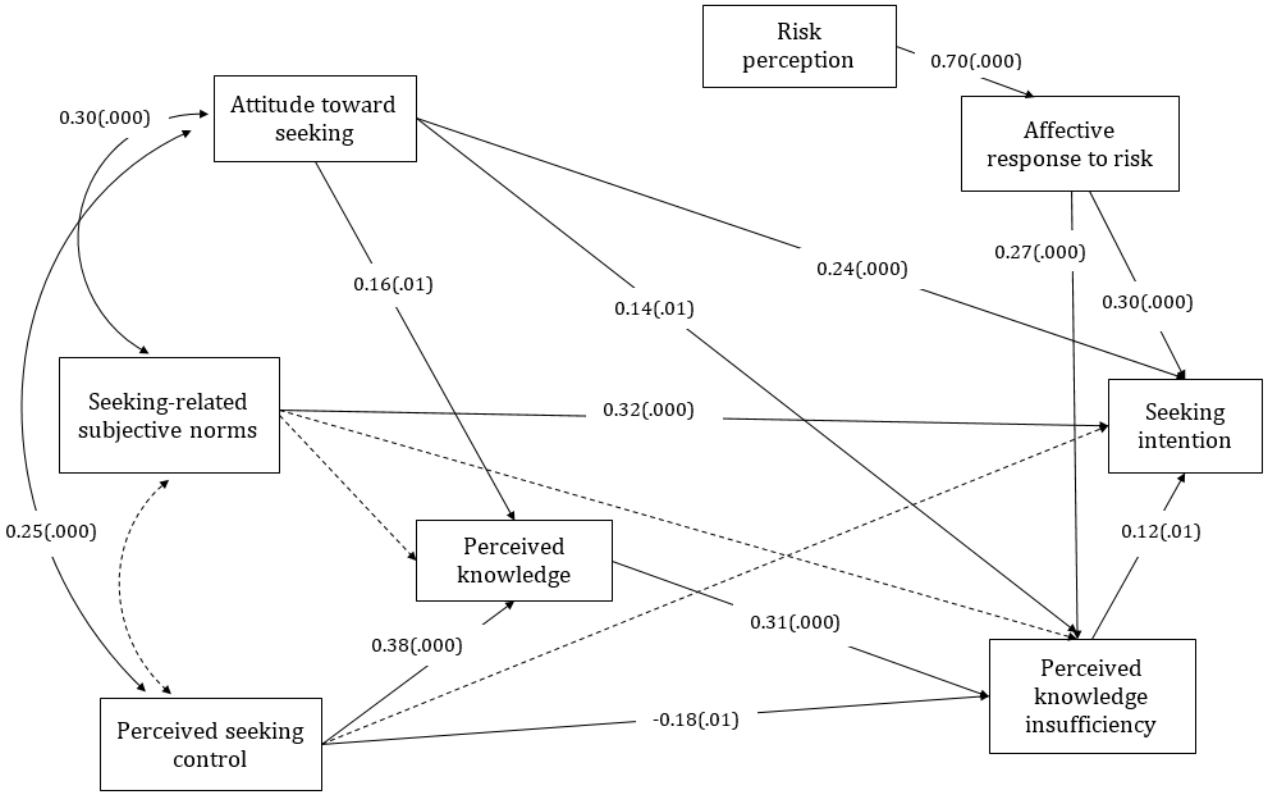
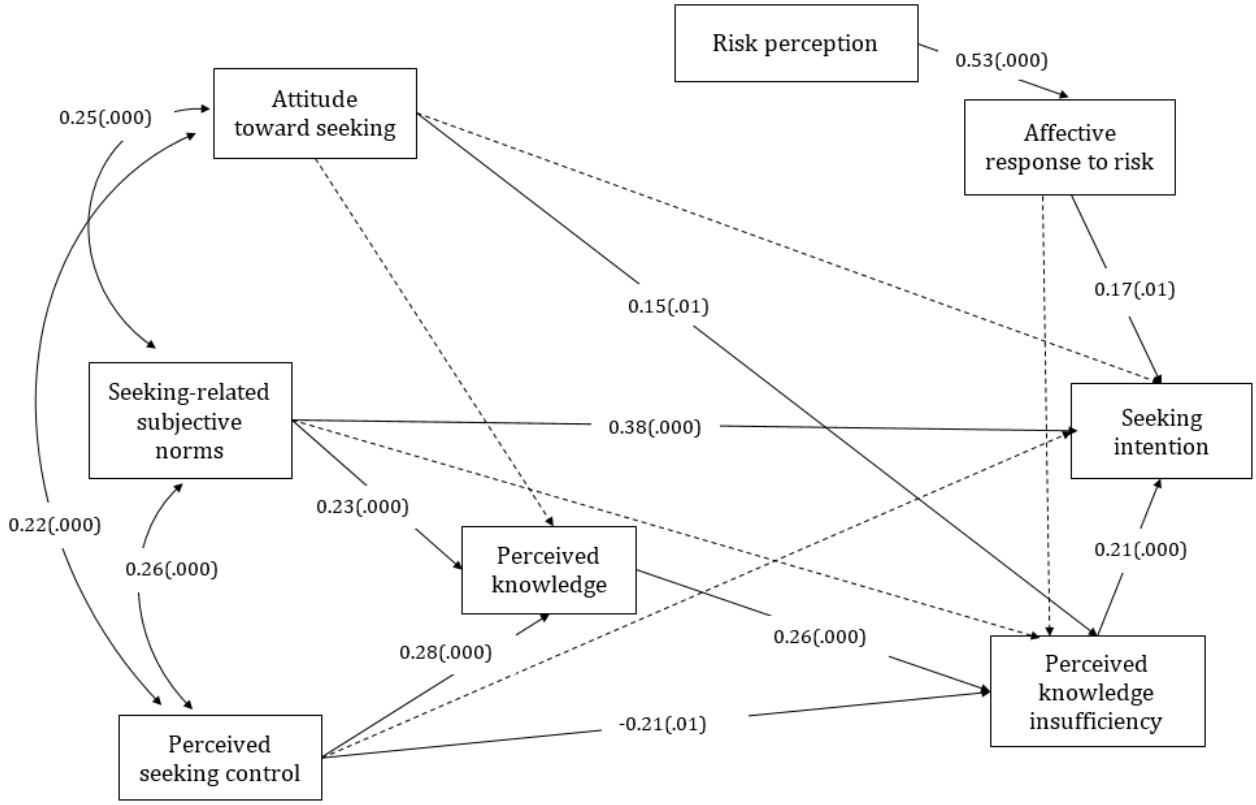


Figure 2. The planned risk information seeking model of the US sample. Dashed lines denote hypothesized nonsignificant paths. The model includes effects of control variables, which are not displayed. (.000) represents significant path coefficients at the .001 level and (.01), at the .05 level.



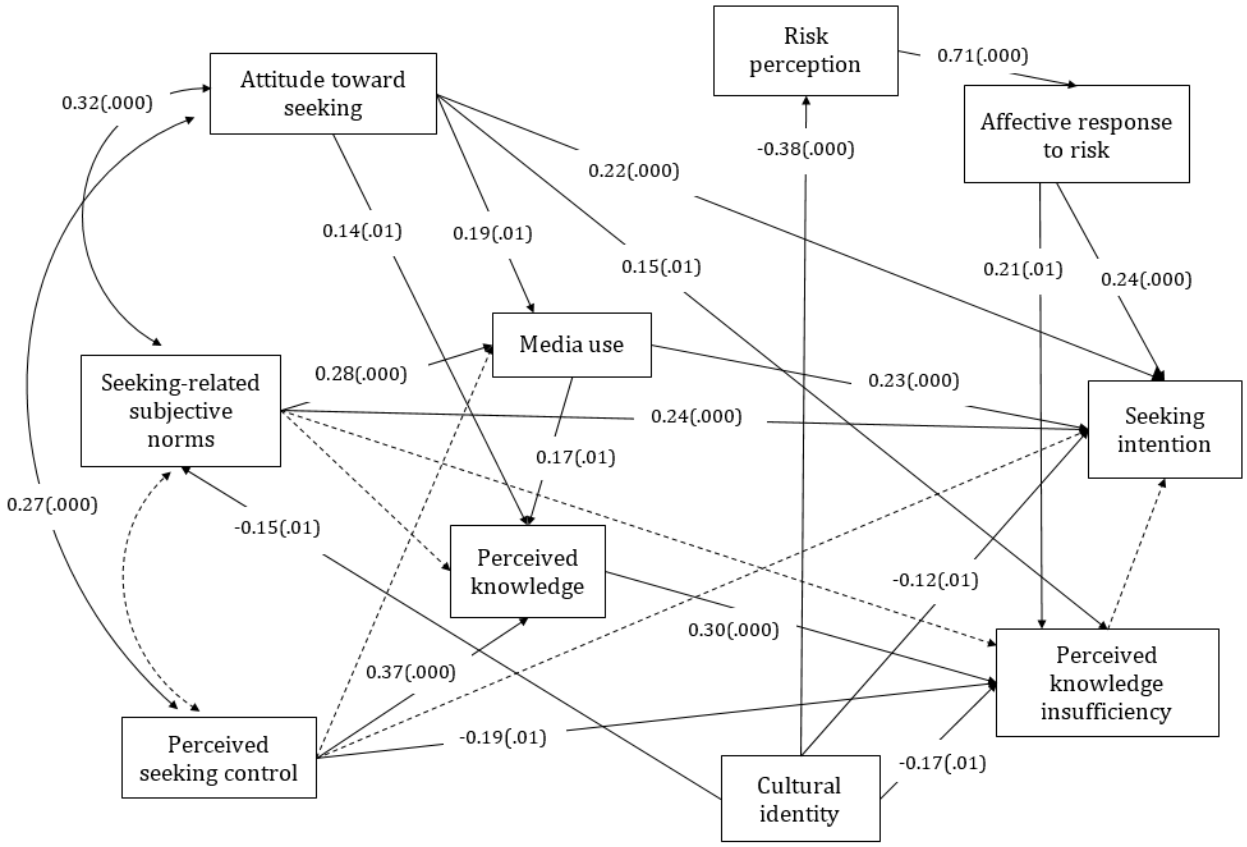
The Extended Planned Risk Information Seeking Model of the Chinese Sample

The Chinese extended model with media use and cultural identity are shown in Figure 3. Attitude toward seeking (β=.22; *P*<.001), seeking-related subjective norms (β=.24; *P*<.001), affective response to risk (β=.24; *P*<.001), media use (β=.23; *P*<.001), and cultural identity (β=−.12; *P*=.02) were significantly associated with information seeking intention. Therefore, H1a and H2 were supported.

In addition, attitude toward seeking (β=.15; *P*=.02) and seeking-related subjective norms (β=.28; *P*<.001) and perceived knowledge (β=.17; *P*=.002) were significantly associated with media use. Cultural identity was significantly associated with perceived knowledge insufficiency (β=−.17; *P*=.003), risk perception (β=−.38; *P*<.001), and seeking-related subjective norms (β=−.15; *P*=.02). These answered RQ2 (What are the relationships between media use and the variables of the PRISM?) and RQ3 (What are the relationships between cultural identity and variables of the PRISM?) in the Chinese sample.

The replicated Chinese PRISM accounted for 35.9% of the variance in information seeking intention, whereas the extended model in the Chinese sample accounted for 41.4% of the variance in information seeking intention, answering RQ4a.

Figure 3. The extended planned risk information seeking model of the Chinese sample. Dashed lines denote hypothesized nonsignificant paths. The model includes effects of control variables, which are not displayed. (.000) represents significant path coefficients at the .001 level and (.01), at the .05 level.



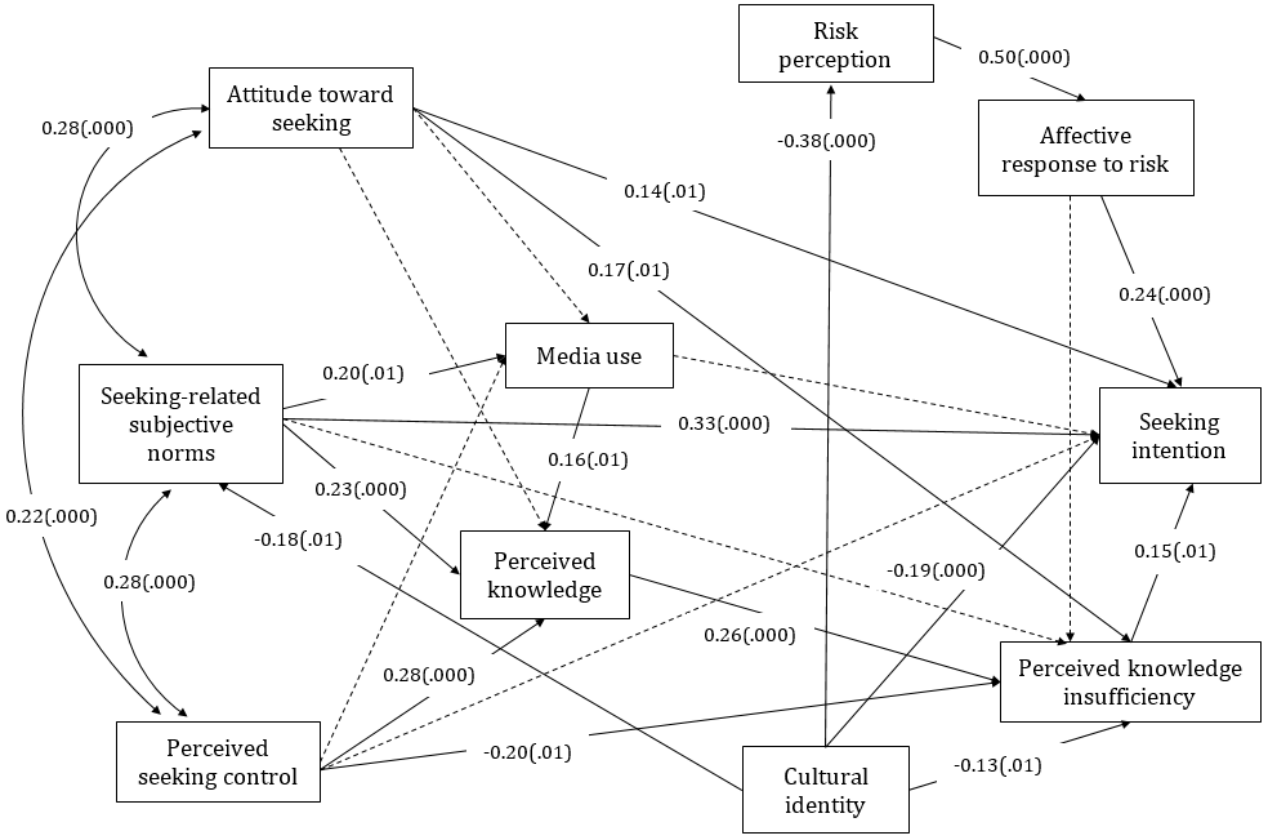
The Extended Planned Risk Information Seeking Model of the US Sample

In the US extended PRISM (see Figure 4), attitude toward seeking (β=.14; *P*=.005), seeking-related subjective norms (β=.35; *P*<.001), perceived knowledge insufficiency (β=.14; *P*=.005), affective response to risk (β=.24; *P*<.001), and cultural identity (β=−.19; *P*<.001) were significantly associated with information-seeking intentions. Thus, H1b and H3 were not supported.

In addition, media use was significantly associated with subjective norms (β=.20; *P*=.001) and perceived knowledge (β=.16; *P*=.004). Perceived knowledge insufficiency (β=−.13; *P*=.02), risk perception (β=−.38; *P*<.001), and seeking-related subjective norms (β=−.18; *P*=.002) were significantly associated with cultural identity. These answered RQ2 (What are the relationships between media use and the variables of the PRISM?) and RQ3 (What are the relationships between cultural identity and variables of the PRISM?) in the US sample.

In the US sample, the PRISM accounted for 30.6% of the variance in information seeking intention, whereas the extended model accounted for 40.3% of the variance in information seeking intention. So RQ4b, which asked whether the extended PRISM accounted for more variance in seeking intention than the PRISM in the US sample, was answered.

Figure 4. The extended planned risk information seeking model of the US sample. Dashed lines denote hypothesized nonsignificant paths. The model includes effects of control variables, which are not displayed. (.000) represents significant path coefficients at the .001 level and (.01), at the .05 level.



Discussion

Principal Findings

In this study, we tested the PRISM model in a sample of US and Chinese college students in the context of mental health. We then added additional variables to the model and compared results across cultures. In terms of the model, the original PRISM model was not a great fit with our data in either samples. Consistent with the results of Kahlor [22], we found that attitude toward seeking and seeking-related subjective norms significantly predicted intention to seek mental health information in both samples. In addition, attitude toward seeking and perceived seeking control significantly predicted perceived knowledge insufficiency. Perceived seeking control influenced seeking intention indirectly by its effects on perceived knowledge and perceived knowledge insufficiency in both samples. Some studies have found a significant role of perceived control in predicting intent to seek help regarding mental health in China [46]. However, perceived control of searching for mental health information does not require professional evaluations, accessibility, cost, and time, which indicates that for personal risks, such as mental health, perceived seeking control may not directly be related to information seeking intention. Risk perception was found to have a direct effect on affective response to risk in all of our models, consistent with the path in the PRISM [22]. Affective response to risk significantly predicted seeking intention and perceived knowledge insufficiency in both samples.

The extended PRISM that we tested included 2 additional variables, ie, media use and cultural identity, and had a better model fit than the PRISM in both the US and Chinese samples. An important finding in this study is the role of media use in PRISM. Our finding is inconsistent with Ho et al's [27] study. Their study demonstrated that media use impacted risk perception and impacted affective response directly and seeking intention indirectly regarding climate change among a Singapore population; however, among our Chinese sample, we found that media use was directly associated with intention to seek mental health information. Mental health as a personal threat could produce a stronger emotional response than impersonal risks, such as climate change [27]. People in China who pay more attention to this issue may have a greater intention to explore more about this personal risk, as they have limited knowledge of professional mental health services and counseling. Another explanation is that people who pay more attention to mental health information have the need for mental health information, so they intend to seek more information regarding mental health [47]. In the Chinese sample, media use was influenced by attitude toward seeking and subjective norms and impacted perceived knowledge, which is consistent with the findings of Ho et al [27]. In the US sample, media use was influenced by subjective norms and had influences on perceived knowledge. We did not find a direct effect of media use on seeking intention in the US sample, which is consistent with the study by Ho et al [27].

Another contribution of this study was that cultural identity significantly predicted seeking intention, perceived knowledge

insufficiency, risk perception, and subjective norms related to information seeking in both samples, which contributes to the extended PRISM in predicting personal risk. However, cultural identity negatively predicted seeking intention in both samples, which indicates that lower levels of identification of the cultural group is associated with higher levels of seeking intention toward mental health information seeking. We predicted that cultural identity would have opposite effects on the Chinese and US participants' information seeking intentions based on the assumption that mental health problems are stigmatized in China while less so in the United States. However, the results showed that people who are less identified with the cultural values regarding mental health issues in their cultures are more likely to seek mental health information, which suggests that mental health issues are stigmatized to some extent in both samples.

Cultural identity is positively related to self-esteem and well-being [48]. If the respondent scored lower on cultural identity, he or she is more likely to have an unclear group identity. This finding reveals that for an issue, which was perceived as stigmatized in both individualistic and collectivistic cultures [6,10], a lower level of cultural identity may act as an unclear group identity to reduce the constraints of a group culture, which could have a direct influence on mental health information seeking. Previous research has found that an unclear cultural identity could reflect a high level of normlessness, which means no clear norms guidance [37]. We found that lower levels of cultural identity were associated with high levels of subjective norms related to mental health information seeking, which means that perceptions of the people with a lower cultural identity about others' expectations on mental health information seeking are inconsistent with the majority of people in the society. Several studies have shown that culture has an impact on risk perceptions among different culture groups [38,49]. On the basis of our results, people with unclear or dissimilar group identities tend to have higher risk perception about mental health information seeking.

Overall, we found that media use and cultural identity can be two variables useful for predicting seeking intentions regarding personal risks such as mental health. With 2 additional variables added in the PRISM, variance accounted for in seeking intentions by the models increased in both samples. Owing to the specificity of the context, the 2 samples with different cultural backgrounds have distinguished results in some perspectives, which again emphasize the importance of culture. We believe that the traditional thought of mental health in China played an important role in explaining these findings as mental health is still a stigmatized issue in China. In addition, the mental health counseling system is not as thoroughly developed in China [50] as in the United States, making this topic less familiar among our Chinese sample. Such differences in perceptions of mental health may directly alter the disease progression. As a chronic disease, early interventional treatment of mental health and the access to additional information during long-term treatment can significantly affect the prognosis of patients, which illustrate the importance of correctly recognizing a mental health issue at an early stage.

These findings can be used in campaigns promoting the seeking of mental health information and can help researchers and practitioners understand the process of personal risk information seeking, especially when the topic has a strong cultural context. The findings also have theoretical implications. Some of the paths are inconsistent from past research regarding personal cancer risk using PRISM [29], and we found that additional variables may provide a better fit for the models based on our topic of mental health.

Limitations and Future Studies

This study is not without limitation. First, we used a convenience sample for both populations. The results of using the sample of college students may not generalize to other subgroups, although it does allow us to assess relationships between key constructs. Second, this study used cross-sectional data and only measured variables at a specific time point without the ability to decide causality. Third, due to a relatively small sample size, we chose to conduct a path analysis. Further studies could conduct a structural equation model. In addition, as our context is a personal risk and often a stigmatized issue, future studies could use a more general concept to examine the role of media use and cultural identity. Moreover, the spiral role of media use should be explored in the future as media use can also influence risk perception and other variables such as attitudes and subjective norms.

Conclusions

Overall, this study extended the PRISM to include important aspects that could vary based on culture, including cultural identity and media use. We found that both cultural identity and media use were associated with information-seeking intentions regarding mental health topics and including the variables in the PRISM allowed the model to account for additional variance in information-seeking intentions. These results can help researchers and health practitioners as they continue to grapple with sensitive health and risk issues, such as mental health. Potential patients may benefit from the findings, changing the progression of the disease and how it is treated, leading to more appropriate treatment solutions.

Conflicts of Interest

None declared.

References

1. Substance Abuse and Mental Health Services Administration. 2015 Sep. Behavioral Health Trends in the United States: Results from the 2014 National Survey on Drug Use and Health URL: http://www.samhsa.gov/data/sites/default/files/NSDUH-FRR1-2014/NSDUH-FRR1-2014.pdf [accessed 2019-06-16]

2. Xiang Y, Yu X, Sartorius N, Ungvari GS, Chiu HF. Mental health in China: challenges and progress. Lancet 2012 Nov 17;380(9855):1715-1716. [doi: 10.1016/S0140-6736(11)60893-3] [Medline: 23158236]
3. Furnham A, Henley S. Lay beliefs about overcoming psychological problems. J Soc Clin Psychol 1988 Sep;6(3-4):423-438. [doi: 10.1521/jscp.1988.6.3-4.423]
4. Granello DH, Granello PF. Defining mental illness: the relationship between college students' beliefs about the definition of mental illness and tolerance. J Col Couns 2000;3(2):100-112. [doi: 10.1002/j.2161-1882.2000.tb00170.x]
5. Yoon E, Jepsen DA. Expectations of and attitudes toward counseling: a comparison of Asian international and US graduate students. Int J Adv Counselling 2008 Apr 3;30(2):116-127. [doi: 10.1007/s10447-008-9050-4]
6. Root MP. Guidelines for facilitating therapy with Asian-American clients. Psychotherapy (Chic) 1985;22(2S):349-356. [doi: 10.1037/h0085514]
7. Zane N, Mak W. Major approaches to the measurement of acculturation among ethnic minority populations: a content analysis and an alternative empirical strategy. In: Chun KM, Balls Organista P, Marín G, editors. Acculturation: Advances in Theory, Measurement, and Applied Research. Washington, DC: American Psychological Association; 2003:39-60.
8. Goldberg D, Bridges K. Somatic presentations of psychiatric illness in primary care setting. J Psychosom Res 1988 Jan;32(2):137-144. [doi: 10.1016/0022-3999(88)90048-7]
9. Cox TH, Lobel SA, McLeod PL. Effects of ethnic group cultural differences on cooperative and competitive behavior on a group task. Acad Manag 1991 Dec;34(4):827-847. [doi: 10.5465/256391]
10. Rao H, Mahadevappa H, Pillay P, Sessay M, Abraham A, Luty J. A study of stigmatized attitudes towards people with mental health problems among health professionals. J Psychiatr Ment Health Nurs 2009 Apr;16(3):279-284. [doi: 10.1111/j.1365-2850.2008.01369.x] [Medline: 19291157]
11. International Telecommunication Union. 2015. ITU releases 2015 ICT figures: Statistics confirm ICT revolution of the past 15 years URL: http://www.itu.int/net/pressoffice/press_releases/2015/17.aspx#.VsSsE_krK71 [accessed 2017-08-16]
12. Pew Internet. 2006. Part 1. 113 Million Internet Users Seek Health Information Online URL: http://www.pewinternet.org/2006/10/29/part-1-113-million-internet-users-seek-health-information-online/ [accessed 2017-08-16]
13. Liu J, Liu Z, Zhang Z, Dong S, Zhen Z, Man L, et al. Internet usage for health information by patients with epilepsy in China. Seizure 2013 Nov;22(9):787-790 [FREE Full text] [doi: 10.1016/j.seizure.2013.06.007] [Medline: 23831148]
14. Powell J, Clarke A. Internet information-seeking in mental health: population survey. Br J Psychiatry 2006 Sep;189:273-277 [FREE Full text] [doi: 10.1192/bjp.bp.105.017319] [Medline: 16946364]
15. Horgan A, Sweeney J. Young students' use of the internet for mental health information and support. J Psychiatr Ment Health Nurs 2010 Mar;17(2):117-123. [doi: 10.1111/j.1365-2850.2009.01497.x] [Medline: 20465756]
16. National Alliance on Mental Illness. Mental health facts in America URL: http://www.nami.org/NAMI/media/NAMI-Media/Infographics/GeneralMHFacts.pdf [accessed 2017-02-02]
17. National Alliance on Mental Illness. 2016. Starting the conversation: College and your mental health URL: http://nami.org/collegeguide/download [accessed 2019-06-16]
18. Li H, Li W, Liu Q, Zhao A, Prevatt F, Yang J. Variables predicting the mental health status of Chinese college students. Asian J Psychiatr 2008 Dec;1(2):37-41. [doi: 10.1016/j.ajp.2008.09.003]
19. Mackenzie S, Wiegel JR, Mundt M, Brown D, Saewyc E, Heiligenstein E, et al. Depression and suicide ideation among students accessing campus health care. Am J Orthopsychiatry 2011;81(1):101-107. [doi: 10.1111/j.1939-0025.2010.01077.x]
20. Rideout VJ, Foehr UG, Roberts DF. Education Resources Information Center. 2010. Generation M2: Media in the Lives of 8- to 18-Year-Olds URL: https://eric.ed.gov/?id=ED527859 [accessed 2019-06-16]
21. Lambert SD, Loiselle CG. Health information seeking behavior. Qual Health Res 2007 Oct;17(8):1006-1019. [doi: 10.1177/1049732307305199] [Medline: 17928475]
22. Kahlor L. PRISM: a planned risk information seeking model. Health Commun 2010 Jun;25(4):345-356. [doi: 10.1080/10410231003775172] [Medline: 20512716]
23. Ajzen I. From intentions to actions: a theory of planned behavior. In: Kuhl J, Beckmann J, editors. Action-Control: From Cognition to Behavior. Heidelberg: Springer; 1985:11-39.
24. Griffin RJ, Dunwoody S, Neuwirth K. Proposed model of the relationship of risk information seeking and processing to the development of preventive behaviors. Environ Res 1999 Feb;80(2 Pt 2):S230-S245. [doi: 10.1006/enrs.1998.3940] [Medline: 10092438]
25. Kahlor LA. An augmented risk information seeking model: the case of global warming. Media Psychology 2007 Sep 28;10(3):414-435. [doi: 10.1080/15213260701532971]
26. Yang ZJ, Kahlor L. What, me worry? The role of affect in information seeking and avoidance. Sci Commun 2012 May 24;35(2):189-212. [doi: 10.1177/1075547012441873]
27. Ho SS, Detenber BH, Rosenthal S, Lee EW. Seeking information about climate change: effects of media use in an extended PRISM. Sci Commun 2014 Feb 07;36(3):270-295. [doi: 10.1177/1075547013520238]
28. Willoughby JF, Myrick JG. Does context matter? Examining PRISM as a guiding framework for context-specific health risk information seeking among young adults. J Health Commun 2016 Jun;21(6):696-704. [doi: 10.1080/10810730.2016.1153764] [Medline: 27187057]

29. Hovick SR, Kahlor L, Liang M. Personal cancer knowledge and information seeking through PRISM: the planned risk information seeking model. J Health Commun 2014 Apr;19(4):511-527. [doi: 10.1080/10810730.2013.821556] [Medline: 24433251]
30. Bronfman NC, Cifuentes LA. Risk perception in a developing country: the case of Chile. Risk Anal 2003 Dec;23(6):1271-1285. [doi: 10.1111/j.0272-4332.2003.00400.x]
31. Slater MD. Reinforcing spirals: the mutual influence of media selectivity and media effects and their impact on individual behavior and social identity. Commun Theory 2007;17(3):281-303. [doi: 10.1111/j.1468-2885.2007.00296.x]
32. Zhao X. Personal values and environmental concern in China and the US: the mediating role of informational media use. Commun Monogr 2012 Jun;79(2):137-159. [doi: 10.1080/03637751.2012.672999]
33. Tian Y, Robinson JD. Media use and health information seeking: an empirical test of complementarity theory. Health Commun 2008;23(2):184-190. [doi: 10.1080/10410230801968260] [Medline: 18444004]
34. Ackerson LK, Viswanath K. Media attention and public perceptions of cancer and eastern equine encephalitis. J Community Health 2010 Aug;35(4):409-416. [doi: 10.1007/s10900-010-9257-2] [Medline: 20352480]
35. Mueller NT, Noone A, Luta G, Wallington SF, Huerta EE, Mandelblatt JS, Latin American Cancer Research Coalition. Information channels associated with awareness of human papillomavirus infections and vaccination among Latino immigrants from safety net clinics. J Immigr Minor Health 2012 Feb;14(1):183-188 [FREE Full text] [doi: 10.1007/s10903-011-9501-6] [Medline: 22089978]
36. Slater MD, Jain P. Teens' attention to crime and emergency programs on television as a predictor and mediator of increased risk perceptions regarding alcohol-related injuries. Health Commun 2011 Jan;26(1):94-103 [FREE Full text] [doi: 10.1080/10410236.2011.527625] [Medline: 21240701]
37. Usborne E, Taylor DM. The role of cultural identity clarity for self-concept clarity, self-esteem, and subjective well-being. Pers Soc Psychol Bull 2010 Jul;36(7):883-897. [doi: 10.1177/0146167210372215] [Medline: 20519575]
38. Niu Z, Willoughby JF. Examining cultural identity and media use as predictors of intentions to seek mental health information among Chinese. Asian J Commun 2018 Feb 17;28(4):360-376. [doi: 10.1080/01292986.2018.1441319]
39. Triandis HC, Hofstede G. Cultures and organizations: software of the mind. Adm Sci Q 1993 Mar;38(1):132. [doi: 10.2307/2393257]
40. Hofstede G. The business of international business is culture. Int Bus Rev 1994 Mar;3(1):1-14. [doi: 10.1016/0969-5931(94)90011-6]
41. Cho H, Lee J. Collaborative information seeking in intercultural computer-mediated communication groups: testing the influence of social context using social network analysis. Commun Res 2008 May 19;35(4):548-573. [doi: 10.1177/0093650208315982]
42. Yang ZJ, Kahlor L, Li H. A United States-China comparison of risk information–seeking intentions. Commun Res 2013 Mar 12;41(7):935-960. [doi: 10.1177/0093650213479795]
43. Brossard D, Nisbet MC. Deference to scientific authority among a low information public: understanding US opinion on agricultural biotechnology. Int J Public Opin Res 2006 Mar 13;19(1):24-52. [doi: 10.1093/ijpor/edl003]
44. Browne MW, Cudeck R. Alternative ways of assessing model fit. Sociol Methods Res 2016 Jun 29;21(2):230-258. [doi: 10.1177/0049124192021002005]
45. Kline R. Principles and Practice of Structural Equation Modeling, Fourth Edition. New York: The Guilford Press; 2005.
46. Mo PK, Mak WW. Help-seeking for mental health problems among Chinese. Soc Psychiat Epidemiol 2009 Mar 4;44(8):675-684. [doi: 10.1007/s00127-008-0484-0]
47. Zhao X. Media use and global warming perceptions: a snapshot of the reinforcing spirals. Commun Res 2009 Jul 17;36(5):698-723. [doi: 10.1177/0093650209338911]
48. Campbell JD. Self-esteem and clarity of the self-concept. Pers Soc Psychol Bull 1990;59(3):538-549. [doi: 10.1037/0022-3514.59.3.538]
49. Gierlach E, Belsher B, Beutler L. Cross-cultural differences in risk perceptions of disasters. Risk Anal 2010 Oct;30(10):1539-1549. [doi: 10.1111/j.1539-6924.2010.01451.x] [Medline: 20626692]
50. Chan B, Parker G. Some recommendations to assess depression in Chinese people in Australasia. Aust N Z J Psychiatry 2004 Mar;38(3):141-147. [doi: 10.1080/j.1440-1614.2004.01321.x] [Medline: 14961932]

Abbreviations

PRISM: planned risk information seeking model
RISP: risk information seeking and processing

Edited by G Eysenbach; submitted 09.08.19; peer-reviewed by J Sapp, K Binger; comments to author 30.11.19; revised version received 13.12.19; accepted 26.01.20; published 11.05.20.

Please cite as:
Niu Z, Willoughby JF, Mei J, Li S, Hu P
A Cross-Cultural Comparison of an Extended Planned Risk Information Seeking Model on Mental Health Among College Students: Cross-Sectional Study
J Med Internet Res 2020;22(5):e15817
URL: http://www.jmir.org/2020/5/e15817/
doi:10.2196/15817
PMID:32441654

Original Paper

Mediating Effect of Internet Addiction on the Relationship Between Individualism and Cyberbullying: Cross-Sectional Questionnaire Study

Ibrahim Arpaci[1*], BSc, MSc, PhD; Thabet Abdeljawad[2,3,4], BSc, MSc, PhD; Mustafa Baloğlu[5*], BSc, MSc, PhD; Şahin Kesici[6*], BSc, MSc, PhD; Ibrahim Mahariq[7], BSc, MSc, PhD

[1]Department of Computer Education and Instructional Technology, Tokat Gaziosmanpasa University, Tokat, Turkey

[2]Department of Mathematics and General Sciences, Prince Sultan University, Riyadh, Saudi Arabia

[3]Department of Medical Research, China Medical University, Taichung, Taiwan

[4]Department of Computer Science and Information Engineering, Asia University, Taichung, Taiwan

[5]Department of Special Education, Faculty of Education, Hacettepe University, Ankara, Turkey

[6]Department of Educational Sciences, Necmettin Erbakan University, Konya, Turkey

[7]College of Engineering and Technology, American University of the Middle East, Kuwait

[*]these authors contributed equally

Corresponding Author:
Ibrahim Arpaci, BSc, MSc, PhD
Department of Computer Education and Instructional Technology
Tokat Gaziosmanpasa University
Faculty of Education
Tokat, 60250
Turkey
Phone: 90 356 252 16 16
Email: ibrahimarpaci@gmail.com

Abstract

Background: Among a variety of dynamics that may have effects on internet-related behaviors, cultural orientation is particularly important. Previous studies suggest that individualism is a strong determinant of certain behaviors. In addition, findings suggest that vertical individualism may lead to the development of more tolerance for addiction and aggression on the internet.

Objective: This study aimed to investigate whether vertical individualism has significant positive effects on cyberbullying and internet addiction and whether horizontal individualism has significant negative effects on cyberbullying and internet addiction. A theoretical model was specified to test the relationships among vertical versus horizontal individualism, cyberbullying, and internet addiction.

Methods: A total of 665 college students were selected using a convenience sampling method and willingly participated in the study. Participants' ages ranged from 17 to 19 years (mean 17.94 years, SD 1.12 years). Of the group, 462 were women (462/665, 69.5%), and 203 were men (203/665, 30.5%). Study majors represented were mathematics (113/665, 17%), sciences (102/665, 15.3%), instructional technology (99/665, 14.9%), psychology (98/665, 14.7%), and others (253/665, 38.1%). Self-report instruments were used to measure vertical/horizontal individualism, cyberbullying, and internet addiction.

Results: Results show a significant positive effect of vertical individualism (effect size 0.10) and significant negative effect of horizontal individualism (effect size –0.12) on cyberbullying. In addition, the direct effect of vertical individualism on internet addiction was significant (effect size 0.28), but the direct effect of horizontal individualism was not (effect size –0.05). Internet addiction had a significant direct effect on cyberbullying (effect size 0.39) as well as an intervening effect on the relationship between vertical individualism and cyberbullying. Results also indicate significant gender differences in cultural patterns and internet addiction.

Conclusions: The findings suggest that horizontal and vertical individualism have significant effects on internet addiction. The findings also suggest that vertical individualists are more vulnerable to internet addiction. Further, the findings indicate a significant relationship between internet addiction and cyberbullying.

(J Med Internet Res 2020;22(5):e16210) doi:10.2196/16210

KEYWORDS

vertical individualism; horizontal individualism; cyberbullying; internet addiction

Introduction

Background

Among a variety of dynamics that may have effects on internet-related behaviors, cultural orientation is particularly noteworthy [1-4]. Hofstede [5] identified four key cultural orientations of people, among which individualism versus collectivism is one of the rather more frequently investigated dimensions [5-8]. In his conceptualization, individualism is defined as the tendency to which "an individual is supposed to take care of himself/herself" [9], and collectivism is the tendency "to which an individual remains integrated into a group" [10]. Triandis [11] argued that individualism and collectivism emerge from status-equal (ie, horizontal) versus status-unequal (ie, vertical) relationships and therefore identified horizontal versus vertical individualism and horizontal versus vertical collectivism. Our study focused on the effects of horizontal versus vertical individualism.

Triandis [11] argued that individualists tend to emphasize an autonomous self-concept, whereas collectivists are inclined to consider themselves as a part of the group. Therefore, an independent versus interdependent self is one of the distinctive characteristics of the two [10]. Contrary to collectivists, individualists prioritize personal goals over the group goals. Internal processes such as attitudes predict social behaviors among individualists. However, among collectivists, social behaviors are predicted by subjective norms, obligations, and perceived duties [12]. Individualists tend to drop a relationship when the cost of the relationship exceeds their personal benefits; however, collectivists try maintaining the relationship even if the cost surpasses their personal benefits [13].

Triandis [11] also identified additional personality characteristics that differentiate between horizontal individualism and vertical individualism. He suggested that vertical individualism defines the self as autonomous, different, and unequal in status with others. Competition is one of the key aspects of vertical individualism. On the other hand, horizontal individualism defines the self as autonomous and independent but also equal to the self of others. Floros et al [14] found that internet addicts exhibit higher impulsivity and help-rejecting behaviors, suggesting that character and personality are significant factors in predicting internet addiction.

Brady [15] described cyberbullying as "the use of communication-based technologies including social networking sites to engage in deliberate harassment or intimidation of other individuals or groups of persons using online speech or expression." Contemporary research shows that cyberbullying is an increasingly epidemic problem among children [16] as well as adolescents [17]. Casas et al [18] found that bullying is strongly influenced by personal and contextual factors. For example, they argue that empathy was a significant predictor of cyberbullying. Similarly, Mishna et al [19] suggested two main risk factors for involvement in cyberbullying: the increasing use of technology [20] and the lack of face-to-face interactions associated with social cues [21,22]. In addition, Smith et al [23] found that being a cyber-victim is correlated with internet use. That is, the more intensive use of the internet, the higher the likelihood of cyberbullying. Therefore, we decided to investigate the effect of internet addiction on cyberbullying.

Based on the cognitive-behavioral model of Davis [24], internet addiction is conceptualized as "an impulse control disorder" [25] and found to be related to a wide range of psychosocial complications [26,27], including cyberbullying [28,29]. Internet addiction or problematic internet use [30] is one of the central research areas for college students. In addition, assessment instruments [31] and screening methods [32] have been developed in the area of problematic internet use or internet addiction. However, the potential association between internet addiction and individuals' cultural orientations (ie, individualism) has not been sufficiently investigated. In general, previous studies showed positive associations between individualism and addictive behavior [33,34] or aggressive behavior [35]. Accordingly, we focused on the effect of the relationship between vertical versus horizontal individualism and internet addiction on cyberbullying.

Cyberbullying and internet addiction have been relatively more frequent themes of recent research [36,37]. However, the lack of previous studies on the impact of cultural individualism on cyberbullying or internet addiction calls for an investigation [38]. Therefore, the purpose of this research was to fill the gap in prior studies by studying the role of internet addiction in the association between vertical versus horizontal individualism and cyberbullying. Such an investigation is warranted for the screening, identification, diagnosis, prevention, and treatment of cyberbullying and internet addiction.

Hypotheses

In general, individualists tend to behave autonomously and prioritize their personal preferences [39]. Hooker [40] argues that individualists have a stronger sense of private space and are more likely to prefer loose personal ties [41]. However, previous studies have reported significant individual differences between horizontal individualists and vertical individualists [42-44].

Horizontal individualism is described as "a model of independent self that fosters a propensity to value uniqueness and social equality," whereas vertical individualism describes "an autonomous self that garners gratification through competition and personal achievement" [8,11]. Vertical individualists are particularly concerned with comparing themselves with others and are likely to enjoy "competition, hedonism, and acquiring status through rivalry" [45]. Vertical individualists prefer to accept inequality and acknowledge the importance of status as well as social rank, whereas horizontal

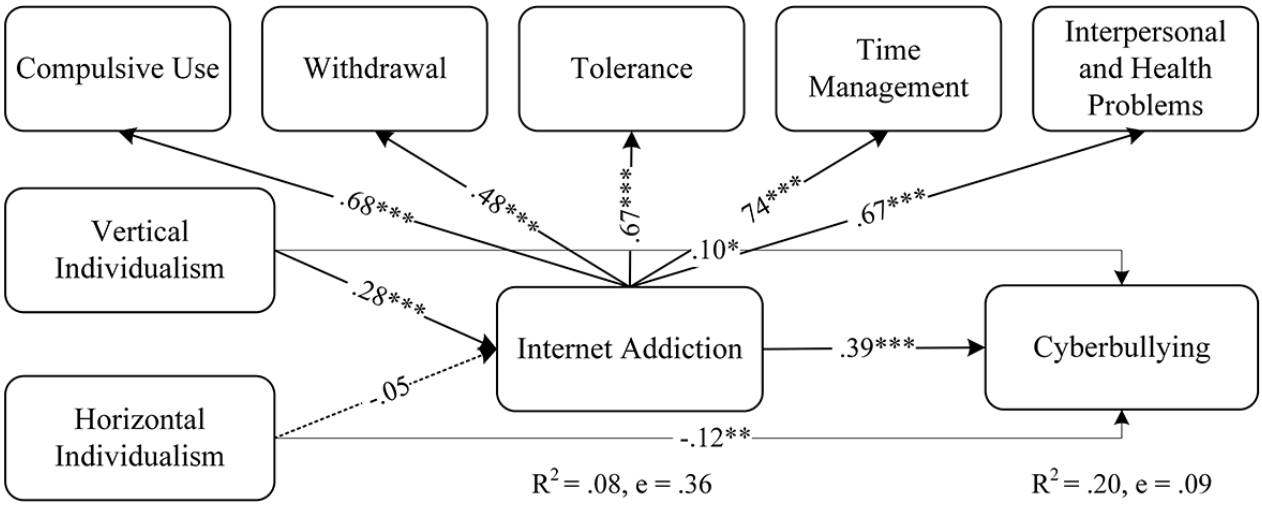
individualists prefer to accept interdependence and equal status for all [11]. We deduced from previous studies that vertical individualists emphasize competition, prestige, hedonism, and status more than horizontal individualists [8].

Previous studies on substance dependence suggested a positive correlation between individualism and addictive behavior [33]. In addition, Bergmüller [35] found that individualism is a strong determinant of aggressive behavior. Ogihara and Uchida [46] found that individualism is negatively related to the number of intimate friends and subjective well-being. These findings suggest that vertical individualism may lead to the development of more tolerance for addiction and aggression. Thus, we theorized that people who score higher for vertical individualism would be more inclined to cyberbullying and internet addiction. We hypothesized that vertical individualism would have a significant positive effect on cyberbullying (H1a) and internet addiction (H1b), whereas horizontal individualism would have a significant negative effect on cyberbullying (H2a) and internet addiction (H2b).

Social repercussions are among the most negative consequences of internet addition [47]. Ko et al [48] suggested that individuals with internet addiction are more likely to have aggressive behaviors. Recent studies also showed significant relationships between internet addition and cyberbullying [28,29]. For example, Gámez-Guadix et al [49] found that cyberbullying was predicted by problematic internet use. You and Lim [29] and Chang et al [50] suggested that internet addiction is associated with cyberbullying. Therefore, we hypothesized that the degree of internet addiction would be positively related to cyberbullying [H3].

According to Hofstede [5], individuals may behave differently depending on their cultural orientations. Therefore, we used horizontal and vertical individualism as extraneous variables in the theoretical model shown in Figure 1. Further, we included internet addiction as an intervening variable between individualism and cyberbullying.

Figure 1. Relationships between vertical-horizontal individualism, cyberbullying, and internet addiction.



*p<.05, **p<.01,***p<.001, Chi-Square = 1429.18, DF = 777, Chi-Square/df = 1.84, GFI = .91, RMSEA = .036

Methods

Participants

A total of 665 freshmen from two state universities in the central part of Turkey who were selected using a convenience sampling method willingly participated in the study. The participants completed an anonymous online survey and received extra course credit for participation. Participants' ages ranged from 17 to 19 years (mean 17.94 years, SD 1.12 years). Of the group, 462 were women (462/665, 69.5%), and 203 were men (203/665, 30.5%). Students from mathematics (113/665, 17%), science (102/665, 15.3%), instructional technology (99/665, 14.9%), psychology (98/665, 14.7%), and other departments (253/665, 38.1%) were represented in the study.

Measures

We used a total of 67 items: 18 items for individualism (10 items for horizontal individualism and 8 items for vertical individualism), 23 items for cyberbullying, and 26 items for internet addiction. All instruments asked participants to rate their level of agreement using a 5-point Likert scale ranging from "strongly disagree" to "strongly agree."

The Individualism-Collectivism Scale

Singelis et al [8] developed the Individualism-Collectivism scale to examine differences in vertical versus horizontal individualism and vertical versus horizontal collectivism. Evidence for the validity and reliability of the scale has been documented [8]. Of the 18 individualism items, 8 items measure vertical individualism, and 10 items measure horizontal individualism (for sample items, see Table 1). Wasti and Erdil [51] adapted the scale into Turkish and reported that the Cronbach α internal consistency coefficients were .67 for horizontal individualism and .73 for vertical individualism. However, in this study, we obtained Cronbach α values of .81 and .82 for horizontal and vertical individualism, respectively.

Table 1. Evidence of the validity and reliability of the measures used in the study.

Construct, Sample item	α[a]	Item-total correlation[b]	Factor loading[b]	Communality[b]	Total variance explained[a]
Cyberbullying					
CB1	N/A	.67	.69	.48	N/A
CB2: I create accounts in web-sites, such as Facebook and Twitter, secretly using others' names	N/A	.83	.84	.71	N/A
CB3	N/A	.84	.86	.74	N/A
CB4	N/A	.89	.90	.82	N/A
CB5	N/A	.86	.88	.78	N/A
CB6	N/A	.87	.89	.79	N/A
CB7	N/A	.78	.81	.65	N/A
CB8	N/A	.90	.92	.84	N/A
CB9	N/A	.87	.89	.78	N/A
CB10	N/A	.87	.89	.79	N/A
CB11	N/A	.83	.84	.71	N/A
CB12	N/A	.84	.86	.74	N/A
CB13	N/A	.88	.90	.80	N/A
CB14	N/A	.86	.87	.76	N/A
CB15	N/A	.74	.75	.57	N/A
CB16	N/A	.76	.77	.60	N/A
CB17	N/A	.89	.91	.82	N/A
CB18	N/A	.74	.76	.57	N/A
CB19	N/A	.84	.86	.74	N/A
CB20	N/A	.85	.87	.75	N/A
CB21	N/A	.88	.90	.81	N/A
CB22	N/A	.82	.84	.70	N/A
CB23	N/A	.72	.74	.55	N/A
Total subscale	.98	N/A	N/A	N/A	69.57
Compulsive use					
CU1: I can't control myself when it comes to the internet	N/A	.78	.78	.61	N/A
CU2	N/A	.79	.75	.54	N/A
CU3	N/A	.77	.82	.67	N/A
CU4	N/A	.80	.40	.49	N/A
CU5	N/A	.78	.79	.62	N/A
Total subscale	.82	N/A	N/A	N/A	58.84
Withdrawal					
W1: If I don't use the internet, I feel uncomfortable	N/A	.83	.82	.68	N/A
W2	N/A	.84	.80	.63	N/A
W3	N/A	.86	.73	.54	N/A
W4	N/A	.83	.81	.66	N/A
W5	N/A	.82	.86	.73	N/A

Construct, Sample item	α[a]	Item-total correlation[b]	Factor loading[b]	Communality[b]	Total variance explained[a]
Total subscale	.86	N/A	N/A	N/A	64.68
Tolerance					
T1: I spend more time on the internet than I expect	N/A	.85	.83	.69	N/A
T2	N/A	.83	.87	.76	N/A
T3	N/A	.82	.89	.78	N/A
T4	N/A	.86	.82	.68	N/A
Total subscale	.88	N/A	N/A	N/A	72.82
Time management					
TM1: I use the internet during my sleeping time		.85	.82	.67	N/A
TM2	N/A	.84	.83	.70	N/A
TM3	N/A	.84	.85	.71	N/A
TM4	N/A	.87	.75	.57	N/A
TM5	N/A	.84	.84	.71	N/A
Total subscale	.88	N/A	N/A	N/A	67.20
Interpersonal and health problems					
P1: I neglect my family because of the internet	N/A	.90	.80	.65	N/A
P2	N/A	.90	.80	.64	N/A
P3	N/A	.90	.74	.55	N/A
P4	N/A	.89	.85	.72	N/A
P5	N/A	.89	.84	.70	N/A
P6	N/A	.89	.85	.73	N/A
P7	N/A	.90	.77	.59	N/A
Total subscale	.91	N/A	N/A	N/A	65.31
Horizontal individualism					
HI1	N/A	.79	.75	.57	N/A
HI2	N/A	.79	.52	.40	N/A
HI3: I often do my own thing	N/A	.80	.62	.44	N/A
HI4	N/A	.78	.61	.52	N/A
HI5: I like my privacy	N/A	.78	.79	.68	N/A
HI6	N/A	.78	.65	.51	N/A
HI7	N/A	.78	.67	.51	N/A
HI8	N/A	.80	.77	.59	N/A
HI9	N/A	.80	.64	.41	N/A
HI10	N/A	.80	.67	.46	N/A
Total subscale	.81	N/A	N/A	N/A	50.89
Vertical individualism					
VI1: Winning is everything	N/A	.80	.84	.79	N/A
VI2	N/A	.81	.90	.83	N/A
VI3	N/A	.81	.46	.42	N/A
VI4	N/A	.80	.81	.70	N/A

Construct, Sample item	α[a]	Item-total correlation[b]	Factor loading[b]	Communality[b]	Total variance explained[a]
VI5	N/A	.81	.87	.81	N/A
VI6: When another person does better than I do, I get tense and aroused	N/A	.80	.83	.76	N/A
VI7	N/A	.80	.81	.72	N/A
VI8	N/A	.81	.86	.81	N/A
Total subscale	.82	N/A	N/A	N/A	72.98

[a]Calculated for the subscale only.

[b]Calculated for the subscale items only.

The Internet Addiction Scale

Internet addiction levels were measured using the Internet Addiction Scale [52]. This 26-item scale includes 5 subscales: "compulsive use," "withdrawal," "tolerance," "time management problems," and "interpersonal and health problems" (for sample items, see Table 1). The scale was adapted into Turkish by Kesici and Sahin [53], where they reported satisfactory reliability and validity properties of the Turkish scale. The Cronbach α coefficient of the total scale was .88, and factor loadings ranged from .44 to .74. Likewise, we calculated a Cronbach α value of .86 for the total scale.

The Cyberbullying Scale

Cyberbullying was measured using 23 items [54]. The authors reported that the Cronbach α coefficient of the single-factor scale was .95. We also found a Cronbach α value of .98 for the scale.

Results

Descriptive Findings

Almost all participants had a smartphone, and approximately two-thirds had a notebook. Further, 543 students (543/665, 81.7%) used technology more than 4 hours a day, and 433 students (433/665, 65.1%) used the internet more than 4 hours a day. Pearson correlation analyses showed a significant correlation between internet use and both vertical individualism (r=.11, *P*=.01) and horizontal individualism (r=.09, *P*=.02). In addition, all subscales of the Internet Addiction Scale were significantly positively correlated with cyberbullying. Table 2 reports the descriptive statistics computed on the study variables.

Table 2. Descriptive statistics, bivariate correlations, and principal component analysis.

Variable	1. Cyberbullying	2A. Internet addiction: compulsive use	2B. Internet addiction: withdrawal	2C. Internet addiction: tolerance	2D. Internet addition: time management	2E. Internet addiction: interpersonal health problems	3. Horizontal individualism	4. Vertical individualism
1. Cyberbullying								
r		.29	.31	.28	.44	.42	–.01	.17
P value		.000	.000	.000	.000	.000	.710	.000
2A. Internet addiction: compulsive use								
r	.29	—[a]	.72	.74	.66	.61	.14	.23
P value	.000	—	.000	.000	.000	.000	.000	.000
2B. Internet addiction: withdrawal								
r	.31	.72	—	.67	.61	.56	.13	.23
P value	.000	.000	—	.000	.000	.000	.001	.000
2C. Internet addiction: tolerance								
r	.28	.74	.67	—	.71	.65	.12	.25
P value	.000	.000	.000	—	.000	.000	.003	.000
2D. Internet addition: time management								
r	.44	.66	.61	.71	—	.84	.06	.23
P value	.000	.000	.000	.000	—	.000	.149	.000
2E. Internet addiction: interpersonal health problems								
r	.42	.61	.56	.65	.84	—	.05	.24
P value	.000	.000	.000	.000	.000	—	.185	.000
3. Horizontal individualism								
r	–.01	.14	.13	.12	.06	.05	—	.40
P value	.710	.000	.001	.003	.149	.185	—	.000
4. Vertical individualism								
r	.17	.23	.23	.25	.23	.24	.40	—
P value	.000	.000	.000	.000	.000	.000	.000	—
Mean	26.71	10.37	9.68	8.22	7.77	10.81	40.01	25.47
SD	8.58	3.24	3.54	2.98	3.32	4.37	5.40	5.84
Minimum-Maximum	24-96	5-20	5-20	4-16	5-20	7-28	10-50	8-40
Skewness (SE .10)	.12	.62	.84	.55	.50	.48	–.71	–.09
Kurtosis (SE .19)	1.03	.19	.25	–.22	.96	.03	.97	.20
KMO[b]	.97	.83	.86	.81	.85	.91	.87	.80
Chi-square								
$\chi 2$	19,202	1108	1453	1371	1667	2872	1749	1907
df	276	19	10	6	10	21	45	28
P value	.000	.000	.000	.000	.000	.000	.000	.000

[a]Not applicable.

[b]Kaiser-Meyer-Olkin

Validity and Reliability

Prior to the analyses, data were checked for the adequacy of factor analysis [55,56]. Table 2 also shows the suitability of the data for factor analysis. An exploratory factor analysis was employed by using principal component extraction to assess the construct validity of the scales. The percentages of total variance explained ranged from 50.89% to 72.98%, which are higher than the acceptable minimum value of .40 [57,58]. Each measurement item has a factor loading above .40 and a communality value above .40 [59,60]. The corrected item-total correlation coefficients ranged from .67 to .90, indicating moderate to high homogeneity. Cronbach α coefficients ranged from .81 to .98, indicating good to very good internal consistency [61]. Validity and reliability results are presented in Table 1.

Common Method Bias

Harman's one-factor test was used to check common method bias [62]. All dependent and independent variables were subjected to the exploratory factor analysis. The factors together accounted for 64.30% of the total variance, while the first factor explained only 24.96%. These findings suggested that common method bias was not a concern in the data set.

Structural Model

It was theorized that internet addiction would serve as an intervening variable between vertical versus horizontal individualism and cyberbullying. Structural equation modeling was conducted via maximum likelihood to test the model. Results show that the structural model produced acceptable fit indices (Table 3).

The direct effects of vertical individualism on cyberbullying (β=.10, critical ratio [CR]=2.31, P=.02) and internet addiction (β=.28, CR=5.81, R^2=.08, P<.001) were positively significant. The direct effect of horizontal individualism on cyberbullying (β=–.12, CR=–2.96, P=.01) was significant, but it was not significant on internet addiction (β=–.05, CR=–1.08, P=.32). Therefore, the null hypotheses for H1a, H1b, and H2a were rejected, but H2b failed to be rejected. The proposed path coefficient between internet addiction and cyberbullying was also positive and significant (β=.39, CR=3.72, P<.001). The effect size in this relationship was R^2=.18. Therefore, the null hypothesis for H3 was also rejected. Figure 1 shows the results of the structural equation modeling analysis, including standardized path coefficients and significance levels along with the R-squared values and respective error terms.

Table 3. Model fit indices of the structural model.

Indices	Model	Acceptable values
χ^2	1429.18	N/A[a]
P value	<.001	.05≤P≤1.00 [63]
χ^2_{df}	1.84	<3 [64]
GFI[b]	.91	≥.90 [65]
AGFI[c]	.89	≥.80 [66]
SRMR[d]	.06	≤.10 [64]
RMR[e]	.05	<.05 [67]
RMSEA[f]	.04	<.08 [65]
NFI[g]	.93	≥.90 [65]
TLI[h]	.96	≥.90 [68]
CFI[i]	.96	≥.90 [69]
IFI[j]	.96	≥.90 [70]

[a]A recommended threshold or acceptable value does not exist.

[b]GFI: goodness of fit index.

[c]AGFI: adjusted goodness of fit index.

[d]SRMR: standardized root mean square residual.

[e]RMR: root mean square residual.

[f]RMSEA: root mean square error of approximation.

[g]NFI: normed fit index.

[h]TLI: Tucker Lewis index.

[i]CFI: comparative fit index.

[j]IFI: incremental fit index.

Mediation Analysis

A 4-step approach was used to test the mediation effect of internet addiction [71,72]. First, the direct effect of vertical individualism was significant (β=.27, SE=.05, CR=5.79, *P*<.001), whereas the direct effect of horizontal individualism was not (β=–.40, SE=.06, CR=–1.04, *P*=.30). Second, the direct effects of vertical individualism (β=.20, SE=.02, CR=4.57, *P*<.001) and horizontal individualism (β=–.13, SE=.03, CR=–3.16, *P*<.001) on cyberbullying were significant. Third, the direct effect of internet addiction on cyberbullying was significant (β=.41, SE=.02, CR=9.92, *P*<.001). Finally, the Sobel test showed that the indirect effect of vertical individualism on cyberbullying via the mediator (ie, internet addiction) was significant (CR=5.87, SE=.06, *P*<.001). These results support partial mediation and indicate that vertical individualism has a significant effect on cyberbullying through internet addiction.

Gender Differences

One-way multivariate analysis of variance was used to investigate gender differences between women (462/665) and men (203/665) on vertical versus horizontal individualism. Results showed a significant difference between women and men, where men (mean 26.45 points, SD 6.12 points) scored significantly higher than women (mean 25.04 points, SD 5.66 points) in vertical individualism ($F_{2,662}$=6.42, *P*=.002, Wilk's λ=.98, partial η^2=.02). However, the effect size of gender on vertical individualism, while statistically significant, was practically minimal (partial eta-squared=.02). Multivariate analysis of variance results also showed a significant difference in internet addiction between women and men ($F_{5,659}$=7.97, *P*<.001, Wilk's λ=.94, partial η^2=.06), where men scored higher than women. Further, independent samples *t*-test results suggested no gender difference in cyberbullying between men and women (t_{663}=–1.32, *P*=.19). It is important that a previous study [30] suggested significant differences in internet addiction among age groups. However, the sample in this current study was comprised of a restricted age group (17-19 years old); therefore, we were not able to investigate the differences among age groups.

Discussion

In this research, we investigated the relationships among vertical versus horizontal individualism, cyberbullying, and internet addiction via a theoretical model and provided valuable implications for mental health professionals and researchers. Arpaci et al [73] suggested that individualists face higher socialization problems and the risk of failure in communicating with others. Consequently, they prefer alternative environments, such as cyberspace. Therefore, they are expected to be more prone to internet addiction. Accordingly, we expected significant relationships among individualism, cyberbullying, and internet addiction. Our results showed a stronger correlation between internet use and vertical individualism than between internet use and horizontal individualism. This suggests that vertical individualists tend to spend more hours on the internet and thereby, are more vulnerable to internet addiction.

We found a significant intervening effect of internet addiction in the association between vertical individualism and cyberbullying. These results support the important intervening role of internet addiction in the relationship between vertical individualism and cyberbullying. Thus, knowing the individual's cultural orientation and level of internet addiction might be helpful in the prevention or treatment of their cyberbullying behaviors.

Self-reliance and uniqueness are some of the relatively more positive characteristics of horizontal individualists [74], while competition and hedonism are more of the destructive characteristics of vertical individualists [45]. Thus, we expected that vertical individualism would positively predict cyberbullying. Our findings support that vertical and horizontal individualism are significant positive and negative predictors of cyberbullying, respectively. In general, those who score higher for individualism foster a propensity to avoid establishing close relationships with others and gradually withdraw from society and social environments [75]. In addition, these individuals do not take the initiative to solve problems but are more likely to evade responsibility. As a result, research has associated individualism with problematic behaviors [35,46]. However, we investigated the specific effects of vertical versus horizontal individualism, which have not been studied previously. Based on our findings, we conclude that those individuals who seek distinctiveness and are especially concerned with comparing themselves to others (ie, vertical individualists) tend to show more cyberbullying tendencies. On the contrary, individuals who are more self-reliant and seek individuality (ie, horizontal individualists) show lower cyberbullying tendencies. However, the reader should also keep in mind that both predictions have relatively small effect sizes, meaning that although the predictions are statistically significant, their effects are minimal. In short, we conclude that vertical individualists tend to develop cyberbullying behaviors depending on their level of internet addiction. The points raised in this study should be considered when prevention or treatment programs are being developed for cyberbullying. Different treatment approaches to cyberbullying should be employed depending upon whether the individual is vertically or horizontally oriented and his or her level of internet addiction.

These results also indicate that vertical, but not horizontal, individualism is a significant predictor of internet addiction. Previous research indicates an association between individualism and addictive behaviors [33,34]; however, our results provide details of such a relationship. We conclude that individualists who prefer gratification through competition and personal achievement are under greater risk of internet addiction, not those who tend to just value uniqueness with equality.

The findings of our research confirm the findings of Casas et al [18] who concluded that cyberbullying is affected by internet addiction. We found that higher internet addiction levels predict higher cyberbullying behaviors. Therefore, it seems reasonable to assume that addicts will show higher levels of cyberbullying. This conclusion is consistent with previous literature [29,50,76-78].

The findings show that vertical individualism has a significant effect on internet addiction. More interestingly, the findings suggest that vertical individualists are more vulnerable to internet addiction. Further, the findings indicate a significant relationship between internet addiction and cyberbullying. Therefore, prevention programs for cyberbullying should take cultural orientations into account.

Although vertical and horizontal individualism were perceived to be dichotomous rather than orthogonal, the correlation analysis results shown in Table 2 suggest a positive correlation between both types of individualism. It is important that both horizontal and vertical individualists focus on an autonomous self-concept; however, horizontal individualists place a strong emphasis on equality in status, whereas vertical individualists accept inequalities [8].

Although this study is highly original, it has several limitations. First, the sample used in the study was comprised of a restricted age group and thus homogenous; therefore, the theoretical model needs further confirmation across different age groups. Crosscultural studies should be conducted in different cultures to improve the external validity of the findings. Second, cultural orientations are not the only predictors of cyberbullying, nor is internet addiction the only mediator. However, based on the literature, this study used a single exogenous factor, vertical versus horizontal individualism, and a single mediator, internet addiction, in the structural model. Many other dispositional or situational factors and mediators would be equally worthwhile to explore in future studies. Further, there may be several equivalent models that can predict the impact of both cultural orientations or internet addiction on cyberbullying. This suggests that the proposed model is certainly supported, but not proven; therefore, further studies should replicate the research model.

In this study, vertical and horizontal individualism were studied as cultural orientations of the participants, and they were measured using an individual-level measurement within a monocultural sample. In the same vein, others studied and operationalized the same orientations at the individual level. For example, Bourgeois [79] argued that vertical and horizontal collectivism-individualism are testable dimensions of culture at the individual level. He investigated values (ie freedom and equality) of vertical-horizontal individualists and collectivists by collecting data from active party members of the Republican and Democratic parties in New Brunswick, Canada. Similarly, Le [80] investigated the relationships among vertical individualism, narcissism, immature love, and ludus based on data collected from 179 undergraduate students at University of California Davis. The results suggested that vertical individualism has a positive significant effect on narcissism and immature love. The findings suggested that vertical individualists were more prone to ludic love style and saw others as fulfillment of wishes and needs.

Finally, previous literature perceived vertical and horizontal individualism to be orthogonal [11]. However, our results indicate a positive correlation between vertical and horizontal individualism, suggesting a more complex relationship. Confirming this, Triandis [11] argued that, along with the horizontal-vertical dimension, there are many other dimensions defining different varieties of collectivism and individualism.

Conflicts of Interest

None declared.

References

1. Chau PYK, Cole M, Massey AP, Montoya-Weiss M, O'Keefe RM. Cultural differences in the online behavior of consumers. Commun. ACM 002 Sep 01;45(10):138-143 [FREE Full text] [doi: 10.1145/570907.570911]
2. Chen L, Nath R. Understanding the underlying factors of Internet addiction across cultures: A comparison study. Electronic Commerce Research and Applications 2016 May;17:38-48. [doi: 10.1016/j.elerap.2016.02.003]
3. Jarvenpaa SL, Leidner DE. An Information Company in Mexico: Extending the Resource-Based View of the Firm to a Developing Country Context. Information Systems Research 1998 Dec;9(4):342-361. [doi: 10.1287/isre.9.4.342]
4. Srite, Karahanna. The Role of Espoused National Cultural Values in Technology Acceptance. MIS Quarterly 2006;30(3):679. [doi: 10.2307/25148745]
5. Hofstede G. Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations (2nd ed.). Thousand Oaks, CA: Sage Publications; 2001:-8039.
6. Hofstede G, Bond MH. The Confucius connection: From cultural roots to economic growth. Organizational Dynamics 1988 Mar;16(4):5-21. [doi: 10.1016/0090-2616(88)90009-5]
7. Hofstede G, Minkov M. Cultures and Organizations: Software of the Mind (3rd ed.). New York, NY: McGraw-Hill; 2010.
8. Singelis TM, Triandis HC, Bhawuk DPS, Gelfand MJ. Horizontal and Vertical Dimensions of Individualism and Collectivism: A Theoretical and Measurement Refinement. Cross-Cultural Research 1995;29(3):240-275. [doi: 10.1177/106939719502900302]
9. Hofstede G. Culture's Consequences: International Differences in Work-related Values. Beverly Hills, CA: Sage Publications; 1980.
10. Markus HR, Kitayama S. Culture and the self: Implications for cognition, emotion, and motivation. Psychological Review 1991;98(2):224-253. [doi: 10.1037/0033-295X.98.2.224]
11. Triandis HC. Individualism and collectivism: New directions in social psychology. Boulder, CO: Westview Press; 1995.

12. Miller JG. Cultural Diversity in the Morality of Caring: Individually Oriented Versus Duty-Based Interpersonal Moral Codes. Cross-Cultural Research 1994;28(1):3-39. [doi: 10.1177/106939719402800101]
13. Kim U, Triandis H, Kagitcibasi C, Yoon G. Individualism and Collectivism: Theoretical and Methodological Issues. New-bury Park, CA: Sage; 1994.
14. Floros G, Siomos K, Antoniadis D, Bozikas V, Hyphantis T, Garyfallos G. Examining personality factors and character defenses assists in the differentiation between college students with Internet addiction and unaffected controls. Personality and Individual Differences 2015 Nov;86:238-242. [doi: 10.1016/j.paid.2015.05.030]
15. Brady K. Cyberbullying. In: Russo CJ, editor. Encyclopedia of Law and Higher Education. Thousand Oaks, CA: Sage; 2010:113.
16. Mishna F, McInroy LB, Lacombe-Duncan A, Bhole P, Van Wert M, Schwan K, et al. Prevalence, Motivations, and Social, Mental Health and Health Consequences of Cyberbullying Among School-Aged Children and Youth: Protocol of a Longitudinal and Multi-Perspective Mixed Method Study. JMIR Res Protoc 2016 May 24;5(2):e83 [FREE Full text] [doi: 10.2196/resprot.5292] [Medline: 27220556]
17. Vismara MFM, Toaff J, Pulvirenti G, Settanni C, Colao E, Lavano SM, et al. Internet Use and Access, Behavior, Cyberbullying, and Grooming: Results of an Investigative Whole City Survey of Adolescents. Interact J Med Res 2017 Aug 29;6(2):e9. [doi: 10.2196/ijmr.6231]
18. Casas JA, Del Rey R, Ortega-Ruiz R. Bullying and cyberbullying: Convergent and divergent predictor variables. Computers in Human Behavior 2013 May;29(3):580-587. [doi: 10.1016/j.chb.2012.11.015]
19. Mishna F, Khoury-Kassabri M, Gadalla T, Daciuk J. Risk factors for involvement in cyber bullying: Victims, bullies and bully–victims. Children and Youth Services Review 2012 Jan;34(1):63-70. [doi: 10.1016/j.childyouth.2011.08.032]
20. Palfrey J, Gasser U. Born Digital: Understanding The First Generation Of Digital Natives. New York, NY: Basic Books; 2008.
21. Ang RP. Adolescent cyberbullying: A review of characteristics, prevention and intervention strategies. Aggression and Violent Behavior 2015 Nov;25:35-42. [doi: 10.1016/j.avb.2015.07.011]
22. Ang RP, Goh DH. Cyberbullying Among Adolescents: The Role of Affective and Cognitive Empathy, and Gender. Child Psychiatry Hum Dev 2010 Mar 18;41(4):387-397. [doi: 10.1007/s10578-010-0176-3]
23. Smith PK, Mahdavi J, Carvalho M, Fisher S, Russell S, Tippett N. Cyberbullying: its nature and impact in secondary school pupils. J Child Psychol Psychiatry 2008 Apr;49(4):376-385. [doi: 10.1111/j.1469-7610.2007.01846.x] [Medline: 18363945]
24. Davis R. A cognitive-behavioral model of pathological Internet use. Computers in Human Behavior 2001 Mar;17(2):187-195. [doi: 10.1016/S0747-5632(00)00041-8]
25. Young KS. Psychology of Computer Use: XL. Addictive Use of the Internet: A Case That Breaks the Stereotype. Psychol Rep 2016 Dec 06;79(3):899-902. [doi: 10.2466/pr0.1996.79.3.899]
26. Dhir A, Chen S, Nieminen M. Predicting adolescent Internet addiction: The roles of demographics, technology accessibility, unwillingness to communicate and sought Internet gratifications. Computers in Human Behavior 2015 Oct;51:24-33. [doi: 10.1016/j.chb.2015.04.056]
27. Yang S, Liu Y, Wei J. Social capital on mobile SNS addiction. Internet Research 2016 Aug;26(4):982-1000. [doi: 10.1108/intr-01-2015-0010]
28. Gámez-Guadix M, Orue I, Smith PK, Calvete E. Longitudinal and Reciprocal Relations of Cyberbullying With Depression, Substance Use, and Problematic Internet Use Among Adolescents. Journal of Adolescent Health 2013 Oct;53(4):446-452. [doi: 10.1016/j.jadohealth.2013.03.030]
29. You S, Lim SA. Longitudinal predictors of cyberbullying perpetration: Evidence from Korean middle school students. Personality and Individual Differences 2016 Jan;89:172-176. [doi: 10.1016/j.paid.2015.10.019]
30. Karacic S, Oreskovic S. Internet Addiction Through the Phase of Adolescence: A Questionnaire Study. JMIR Ment Health 2017 Apr 03;4(2):e11 [FREE Full text] [doi: 10.2196/mental.5537] [Medline: 28373154]
31. Bischof-Kastner C, Kuntsche E, Wolstein J. Identifying problematic Internet users: development and validation of the Internet Motive Questionnaire for Adolescents (IMQ-A). J Med Internet Res 2014 Oct 09;16(10):e230 [FREE Full text] [doi: 10.2196/jmir.3398] [Medline: 25299174]
32. Faraci P, Craparo G, Messina R, Severino S. Internet Addiction Test (IAT): which is the best factorial solution? J Med Internet Res 2013 Oct 09;15(10):e225 [FREE Full text] [doi: 10.2196/jmir.2935] [Medline: 24184961]
33. Lopez SG, Garza RT, Gonzalez-Blanks AG. Preventing Smoking Among Hispanic Preadolescents. Hispanic Journal of Behavioral Sciences 2012 Apr 29;34(2):323-339. [doi: 10.1177/0739986311435901]
34. Foster DW, Yeung N, Quist MC. The Influence of Individualism and Drinking Identity on Alcohol Problems. Int J Ment Health Addiction 2014 May 21;12(6):747-758. [doi: 10.1007/s11469-014-9505-2]
35. Bergmüller S. The Relationship Between Cultural Individualism-Collectivism and Student Aggression Across 62 Countries. Aggr. Behav 2013 Mar 12;39(3):182-200. [doi: 10.1002/ab.21472]
36. Görzig A. Adolescents' Viewing of Suicide-Related Web Content and Psychological Problems: Differentiating the Roles of Cyberbullying Involvement. Cyberpsychology, Behavior, and Social Networking 2016 Aug;19(8):502-509. [doi: 10.1089/cyber.2015.0419]

37. Veiga Simão A, Ferreira PC, Freire I, Caetano A, Martins M, Vieira C. Adolescent cybervictimization – Who they turn to and their perceived school climate. Journal of Adolescence 2017 Jul;58:12-23. [doi: 10.1016/j.adolescence.2017.04.009]
38. Genta M, Smith P, Ortega R, Brighi A, Giasrini A, Thompson F. Comparative aspects of cyberbullying in Italy, EnglandSpafindings from a DAPHNE project. In Li Q, Cross D, Smith PK. eds. Cyberbullying in the Global Play Ground: Research from International Perspective. Chichester, England: Wiley-Blackwell; 2012:15-31.
39. Bochner S, Hesketh B. Power Distance, Individualism/Collectivism, and Job-Related Attitudes in a Culturally Diverse Work Group. Journal of Cross-Cultural Psychology 2016 Jul 27;25(2):233-257. [doi: 10.1177/0022022194252005]
40. Hooker J. Working Across Cultures. Palo Alto, CA: Stanford University Press; 2003.
41. Masuda A, Poelmans S, Allen T, Spector P, Lapierre L, Cooper C. Flexible work arrangements availability and their relationship with work-to-family conflict, job satisfaction, and turnover intentions: a comparison of three country clusters. App Psyc 2012;61(1):a. [doi: 10.1111/j.1464-0597.2011.00453.x]
42. Triandis HC, Singelis TM. Training to recognize individual differences in collectivism and individualism within culture. International Journal of Intercultural Relations 1998 Feb;22(1):35-47. [doi: 10.1016/S0147-1767(97)00034-5]
43. Schwartz S. In: U. Kim, H. C. Triandis, Ç. Kâğitçibaşi, S.-C. Choi, & G. Yoon , editor. Beyond individualismcollectivism: New cultural dimensions of values. In Kim U, Triandis HC, Kagitcibasi C, Choi SC, Yoon G. eds. Newbury Park, CA: Sage Publications; 1994:85-122.
44. Hui C. Measurement of individualism-collectivism. Journal of Research in Personality 1988 Mar;22(1):17-36. [doi: 10.1016/0092-6566(88)90022-0]
45. Park H, Rehg MT, Lee D. The Influence of Confucian Ethics and Collectivism on Whistleblowing Intentions: A Study of South Korean Public Employees. J Bus Ethics 2005 Jun;58(4):387-403. [doi: 10.1007/s10551-004-5366-0]
46. Ogihara Y, Uchida Y. Does individualism bring happiness? Negative effects of individualism on interpersonal relationships and happiness. Front Psychol 2014 Mar;5:135 [FREE Full text] [doi: 10.3389/fpsyg.2014.00135] [Medline: 24634663]
47. Douglas AC, Mills JE, Niang M, Stepchenkova S, Byun S, Ruffini C, et al. Internet addiction: Meta-synthesis of qualitative research for the decade 1996–2006. Computers in Human Behavior 2008 Sep;24(6):3027-3044. [doi: 10.1016/j.chb.2008.05.009]
48. Ko C, Yen J, Liu S, Huang C, Yen C. The Associations Between Aggressive Behaviors and Internet Addiction and Online Activities in Adolescents. Journal of Adolescent Health 2009 Jun;44(6):598-605. [doi: 10.1016/j.jadohealth.2008.11.011]
49. Gámez-Guadix M, Gini G, Calvete E. Stability of cyberbullying victimization among adolescents: Prevalence and association with bully–victim status and psychosocial adjustment. Computers in Human Behavior 2015 Dec;53:140-148. [doi: 10.1016/j.chb.2015.07.007]
50. Chang F, Chiu C, Miao N, Chen P, Lee C, Chiang J, et al. The relationship between parental mediation and Internet addiction among adolescents, and the association with cyberbullying and depression. Comprehensive Psychiatry 2015 Feb;57:21-28. [doi: 10.1016/j.comppsych.2014.11.013]
51. Wasti A, Erdil E. Measurement of individualism and collectivism: the validation of the Self-Construal Scale and INDCOL in Turkish. Manag Res J 2007;7(2):39-66.
52. Chen S, Weng L, Su Y, Wu H, Yang P. Development of a Chinese Internet addiction scale and its psychometric study. Chin J Psych 2003;45(3):279-294.
53. Kesici S, Sahin I. Turkish Adaptation Study of Internet Addiction Scale. Cyberpsychology, Behavior, and Social Networking 2010 Apr;13(2):185-189. [doi: 10.1089/cyber.2009.0067]
54. Arıcak O, Kınay H, Tanrıkulu T. Siber Zorbalık Ölçeğinin ilk psikometrik bulguları. Hasan Ali Yücel Egt Fak Derg 2012;9(1):101-114.
55. Kaiser HF. A second generation little jiffy. Psychometrika 1970 Dec;35(4):401-415. [doi: 10.1007/bf02291817]
56. Bartlett MS. The Effect of Standardization on a χ 2 Approximation in Factor Analysis. Biometrika 1951 Dec;38(3/4):337. [doi: 10.2307/2332580]
57. Scherer RF, Luther DC, Wiebe FA, Adams JS. Dimensionality of Coping: Factor Stability Using the Ways of Coping Questionnaire. Psychol Rep 2016 Aug 31;62(3):763-770. [doi: 10.2466/pr0.1988.62.3.763]
58. Arpaci I. Understanding and predicting students' intention to use mobile cloud storage services. Computers in Human Behavior 2016 May;58:150-157. [doi: 10.1016/j.chb.2015.12.067]
59. Field A. Discovering Statistics Using IBM SPSS Statistics. Thousand Oaks, CA: Sage Publications Ltd; 2013.
60. Arpaci I. Antecedents and consequences of cloud computing adoption in education to achieve knowledge management. Computers in Human Behavior 2017 May;70:382-390. [doi: 10.1016/j.chb.2017.01.024]
61. Creswell J. Mixed method designs. In: Creswell JW. , editor. Educational research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research. Upper Saddle River, NJ: Prentice Hall; 2002:146-166.
62. Podsakoff PM, MacKenzie SB, Lee J, Podsakoff NP. Common method biases in behavioral research: a critical review of the literature and recommended remedies. J Appl Psychol 2003 Oct;88(5):879-903. [doi: 10.1037/0021-9010.88.5.879] [Medline: 14516251]
63. Hoyle R. Structural Equation Modeling: Concepts, Issues, And Applications. Thousand Oaks, CA: Sage Publications, Inc; 1995.
64. Kline R. Principles and Practice of Structural Equation Modeling. New York, NY: The Guilford Press; 2005.

65. Hair J, Tatham R, Anderson R, Black C. Multivariate Data Analysis. Upper Saddle River, NJ: Pearson Prentice Hall; 2006.
66. Marsh HW, Balla JR, McDonald RP. Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. Psychological Bulletin 1988;103(3):391-410. [doi: 10.1037/0033-2909.103.3.391]
67. McDonald RP, Ho MR. Principles and practice in reporting structural equation analyses. Psychological Methods 2002;7(1):64-82. [doi: 10.1037/1082-989X.7.1.64]
68. Bentler PM, Bonett DG. Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin 1980;88(3):588-606. [doi: 10.1037/0033-2909.88.3.588]
69. Bentler PM. Comparative fit indexes in structural models. Psychological Bulletin 1990;107(2):238-246. [doi: 10.1037/0033-2909.107.2.238]
70. Bollen K. Structural Equations with Latent Variables. New York, NY: Wiley; 1989.
71. Sobel ME. Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models. Sociological Methodology 1982;13:290. [doi: 10.2307/270723]
72. Hayes A. Introduction To Mediation, Moderation, And Conditional Process Analysis, First Edition: A Regression-based Approach (methodology In The Social Sciences). New York, NY: The Guilford Press; 2020.
73. Arpaci I, Kesici , Baloğlu M. Individualism and internet addiction: the mediating role of psychological needs. Internet Research 2018 Apr 04;28(2):293-314. [doi: 10.1108/intr-11-2016-0353]
74. Hofstede G. Cultures and Organizations. Intercultural Cooperation and its Importance for Survival. Software of the Mind. London: McGraw-Hill; 1991.
75. Triandis HC. Individualism-Collectivism and Personality. J Personality 2001 Dec;69(6):907-924. [doi: 10.1111/1467-6494.696169]
76. Aricak OT, Ozbay A. Investigation of the relationship between cyberbullying, cybervictimization, alexithymia and anger expression styles among adolescents. Computers in Human Behavior 2016 Feb;55:278-285. [doi: 10.1016/j.chb.2015.09.015]
77. Floros GD, Siomos KE, Fisoun V, Dafouli E, Geroukalis D. Adolescent online cyberbullying in Greece: the impact of parental online security practices, bonding, and online impulsiveness. J Sch Health 2013 Jun;83(6):445-453. [doi: 10.1111/josh.12049] [Medline: 23586890]
78. Jung Y, Leventhal B, Kim YS, Park TW, Lee S, Lee M, et al. Cyberbullying, problematic internet use, and psychopathologic symptoms among Korean youth. Yonsei Med J 2014 May;55(3):826-830 [FREE Full text] [doi: 10.3349/ymj.2014.55.3.826] [Medline: 24719154]
79. Bourgeois, DY. Dissertation.: Electronic Theses and Dissertations. 64; 2002. The politics and values of individualists and collectivists: A cross-cultural comparison URL: http://digitalcommons.library.umaine.edu/etd/64 [accessed 2020-05-04]
80. Le TN. Narcissism and immature love as mediators of vertical individualism and ludic love style. Journal of Social and Personal Relationships 2016 Jun 30;22(4):543-560. [doi: 10.1177/0265407505054522]

Abbreviations

AGFI: adjusted goodness of fit index.
CFI: comparative fit index.
GFI: goodness of fit index.
IFI: incremental fit index.
NFI: normed fit index.
RMR: root mean square residual.
RMSEA: root mean square error of approximation.
SRMR: standardized root mean square residual.
TLI: Tucker Lewis index.

Edited by G Eysenbach; submitted 11.09.19; peer-reviewed by K Karataş, T Cahill; comments to author 23.10.19; revised version received 07.12.19; accepted 24.03.20; published 28.05.20.

Please cite as:
Arpaci I, Abdeljawad T, Baloğlu M, Kesici Ş, Mahariq I
Mediating Effect of Internet Addiction on the Relationship Between Individualism and Cyberbullying: Cross-Sectional Questionnaire Study
J Med Internet Res 2020;22(5):e16210
URL: http://www.jmir.org/2020/5/e16210/
doi:10.2196/16210
PMID:32463369

Original Paper

Googling Allergy in Ireland: Content Analysis

Catherine King[1], MB; Ciaran Judge[1], MB; Aideen Byrne[2], PhD, FRCPHC; Niall Conlon[1], PhD, FRCPath

[1]Department of Clinical and Laboratory Immunology, St. James's Hospital, Dublin, Ireland
[2]Paediatric Allergy Department, Our Lady's Children's Hospital, Crumlin, Dublin, Ireland

Corresponding Author:
Catherine King, MB
Department of Clinical and Laboratory Immunology
St. James's Hospital
James's Street
Dublin 8
Ireland
Phone: 353 1 410 3000
Email: cathmarionking@gmail.com

Abstract

Background: Internet search engines are increasingly being utilized as the first port of call for medical information by the public. The prevalence of allergies in developed countries has risen steadily over time. There exists significant variability in the quality of health-related information available on the web. Inaccurately diagnosed and mismanaged allergic disease has major downstream effects on patients, general practitioners, and regional allergy services.

Objective: This study aimed to verify whether Ireland has a relatively high rate of web-based allergy-related searches, to establish the proportion of medically accurate web pages encountered by the public, and to compare current search results localized to Dublin, Ireland with urban centers elsewhere.

Methods: Google Trends was used to evaluate regional interest of allergy-related search terms over a 10-year period using terms “allergy,” “allergy test,” “food allergy,” and “food intolerance.” These terms were then inputted into Google search, localizing them to cities in Ireland, the United Kingdom, and the United States. Output for each search was reviewed by two independent clinicians and deemed rational or nonevidence based, as per current best practice guidelines. Searches localized to Dublin were initially completed in 2015 and repeated in 2019 to assess for changes in the quality of search results over time.

Results: Ireland has a persistently high demand for web-based information relating to allergy and ranks first worldwide for “allergy test,” second for “food allergy” and “food intolerance,” and seventh for “allergy” over the specified 10-year timeframe. Results for each of the four subsearches in Dublin (2015) showed that over 60% of websites promoted nonevidence-based diagnostics. A marginal improvement in scientifically robust information was seen in 2019, but results for “allergy test” and “food intolerance” continued to promote alternative testing 57% (8/14) of the time. This strongly contrasted with results localized to Southampton and Rochester, where academic and hospital-affiliated web pages predominantly featured. Government-funded Department of Health websites did not feature in the top five results for Dublin searches “allergy testing,” “food allergy,” or “food intolerance” in either 2015 or 2019.

Conclusions: The Irish public demonstrates a keen interest in seeking allergy-related information on the web. The proportion of evidence-based websites encountered by the Irish public is considerably lower than that encountered by patients in other urban centers. Factors contributing to this are the lack of a specialist register for allergy in Ireland, inadequate funding for allergy centers currently in operation, and insufficient promotion by the health service of their web-based health database, which contains useful patient-oriented information on allergy. Increased funding of clinical allergology services will more meaningfully impact the health of patients if there is a parallel investment by the health service in information and communication technology consultancy to amplify their presence on the web.

(J Med Internet Res 2020;22(5):e16763) doi:10.2196/16763

KEYWORDS

allergy; food allergy; food intolerance; technology; Ireland; immunology

Introduction

Background

The internet has become a major resource for people seeking information in relation to health. Overall, 57% of Irish adults search for health-related information on the web, and this most often applies to younger people and women [1].

The incidence of allergies has risen steadily in developed countries [2]. Food allergies are thought to affect approximately 1% to 3% of the Irish adult population and are a cause of significant public concern [3,4]. Allergic disease has long been a focus of public attention, with a nationwide questionnaire in 2015 demonstrating that 14% of Irish adults self-report allergic conditions (rhinitis, allergic eye disease, and food allergy, excluding allergic asthma) [5]. The diagnosis of allergic disease is centered upon a detailed clinical history and supported by the judicious use of tests to detect allergic sensitization [6]. Allergic sensitization is determined by skin prick testing or detection of allergen-specific immunoglobulin E (sIgE). When these test results conflict or there is a diagnostic doubt, gold standard investigation is advisable in the form of diagnostic challenges. These time-consuming procedures are only available in specialist centers.

The umbrella term of allergy is one that frequently attracts input from a variety of alternative medicine practitioners, naturopaths, homeopaths, and acupuncturists both in terms of diagnostic testing and claims to treatment. Despite the existence of scientifically robust and evidence-based allergy tests, alternative approaches to diagnosis are widely used by the public, with attempts at regulatory control being previously described as "woefully inadequate" [2]. Alternative tests, including food-specific immunoglobulin G testing, kinesiology, hair analysis, Vega testing, and leukocytotoxic tests, are widely available, lack scientific basis and diagnostic rigor, and have been discredited by a variety of governmental, professional, and expert bodies internationally [6-9]. The use of complementary and alternative medicine (CAM) as a treatment modality for allergy has also been increasingly reported, with 37% of people with allergic disease using CAM during the preceding 12 months in a recent European study [10]. In 2018, 60% of surveyed allergists in the United States had patients who encountered adverse reactions from the use of CAM, with 81% of respondents encountering patients who discontinued conventional therapy while using CAM, irrespective of medical advice [11]. Despite this, alternative approaches continue to be advocated by some health care professionals, including registered medical practitioners. The risks of a misdiagnosis of food allergy include inappropriate dietary restrictions and negative quality-of-life consequences, the misattribution of symptoms to allergic diseases resulting in delayed assessments, the inappropriate fear of life-threatening reactions, and direct and indirect financial costs related to these risks [12].

Objectives

The volume of easily accessible information available on the web offers an excellent opportunity to provide helpful, evidence-based information and services to the information-seeking public. In this study, we sought to examine the web-based information sources accessible to members of the Irish public who were seeking information on allergy testing. We first interrogated allergy testing search requests by Irish internet users. We then examined the prominent sites presented when searching for allergy tests in an Irish setting and determined whether they promoted rational or alternative testing approaches. We initially reviewed these search requests in 2015 and repeated the study using identical search terms in 2019 to identify whether the standard of information available to internet users had changed over time. Finally, we compared the 2019 Irish results with similar UK and US populations using identical search terms.

Methods

Google Trends was used to evaluate interest by region of allergy-related search terms over a 10-year period (January 1, 2009, to December 31, 2018) [13]. The search terms used were "allergy" and related subsearches "allergy test," "food allergy," and "food intolerance." The period of trend analysis was purposely predated to the start of 2019 so that our subsequent searches in July 2019 did not interfere with the trend results. Google Trends data provide a list of countries ranked by the relative popularity of the specified search term, as a proportion of total searches in each country.

Each of the search terms was inputted into Google search and the output reviewed. Output webpages were reviewed independently by two clinicians and classified as *rational*, if the services or information offered were based on clinical history and standard sensitization testing, or *alternative*, if the practitioners offered any non–evidence-based approaches, as outlined in the Irish Food Allergy Network and Irish Association of Allergy and Immunology position statement [7]. Results localized to country level (Ireland) were reviewed initially. Then, a comparison of local results from Dublin, Southampton (United Kingdom), and Rochester (United States) was performed. Sponsored advertisements, forums/discussion groups, duplications, and reports on news or weather were excluded from analysis. Analysis was limited to Google search pages one and two.

Results

Trend Analysis of Allergy-Related Searches Worldwide

Data from Google trends indicate that Ireland has a persistently high demand for information regarding allergy. Trends from 2008 to 2019 demonstrate that Ireland ranks seventh in the world in searches for the term allergy (Table 1). Analysis of data for the subsearch term "allergy test" shows that Ireland is ranked first in the world for this particular search over the timeframe, and second in the world for both other subsearch terms "food allergy" and "food intolerance" [13]. These results demonstrate a high demand for web-based information regarding allergies from the Irish population.

Table 1. Google Trends January 1, 2009, to December 31, 2018, worldwide for the specified search terms.

Search term	Worldwide ranking			Ireland ranking
	First place	Second place	Third place	
"Allergy"	United States	Canada	Philippines	7
"Allergy test"	Ireland	Singapore	United States	1
"Food allergy"	United States	Ireland	Australia	2
"Food intolerance"	Malta	Ireland	Australia	2

Analysis of localized Google search engine results for Ireland (search term "allergy Ireland") in 2015 revealed that alternative diagnostic services and information featured highly (Table 2). At that time, 63% (10/16) of included results were from private companies selling non–evidence-based commercial tests, 19% (3/16) were from private medical services with rational testing procedures, and the remaining 19% (3/16) were from organizations offering evidence-based patient information. Regional localization to Dublin and analysis of the subsearch terms also confirmed the prominence of websites endorsing alternative approaches to allergy diagnostics. Furthermore, 69% (11/16) websites listed under an "allergy test Dublin" Google search promoted alternative non–evidence-based approaches to allergy. Use of the search term "food allergy Dublin" provided similar results, with 67% (10/15) of included results relating to alternative health care websites. The disparity was greater again when "food intolerance Dublin" was used as a search term, with just 13% (2/15) of included webpages promoting a rational assessment approach versus 87% (13/15) promoting alternative approaches.

Table 2. Summary of results of localized internet searches for the specified allergy-related search terms.

Search term, by region	Website results, n (%)		
	Non–evidence-based practices	Medical facility, rational testing	Evidence-based guidelines, journals
"**Allergy**"			
Ireland (2015)	10 (63)	3 (19)	3 (19)
Ireland (2019)	1 (9)	5 (46)	5 (46)
Southampton, United Kingdom (2019)	1 (11)	4 (44)	4 (44)
Rochester, United States (2019)	0 (0)	7 (100)	0 (0)
"**Allergy Test**"			
Dublin (2015)	11 (69)	5 (31)[a]	N/A[b]
Dublin (2019)	8 (57)	5 (36)	1 (7)
Southampton, United Kingdom (2019)	1 (8)	8 (67)	3 (25)
Rochester, United States (2019)	0 (0)	7 (70)	3 (30)
"**Food Allergy**"			
Dublin (2015)	10 (67)	5 (33)[a]	N/A
Dublin (2019)	6 (46)	5 (39)	2 (15)
Southampton, United Kingdom (2019)	1 (9)	4 (36)	6 (55)
Rochester, United States (2019)	0 (0)	5 (50)	5 (50)
"**Food Intolerance**"			
Dublin (2015)	13 (87)	2 (13)[a]	N/A
Dublin (2019)	8 (57)	4 (29)	2 (14)
Southampton, United Kingdom (2019)	3 (27)	6 (55)	2 (18)
Rochester, United States (2019)	1 (11)	3 (33)	5 (56)

[a]The 2015 search results for "Allergy test Dublin," "Food allergy Dublin," and "Food intolerance Dublin" were recorded as either alternative or evidence based only.

[b]N/A: not available.

Subsearch Terms Localized to Ireland (2019)

A greater proportion of evidence-based information was noted upon repeating these Google searches in July 2019. Current searches for "allergy Ireland" result in 9% (1/11) from a private company selling non–evidence-based commercial tests, 45% (5/11) from private health care facilities engaged in rational testing, and a further 45% (5/11) from organizations offering evidence-based patient information. However, searches for "Allergy test Dublin" continued to promote alternative approaches to diagnosing allergy, with 57% (8/14) of the results coming from private companies selling non–evidence-based commercial tests. Results for "Food allergy Dublin" also continued to feature a high proportion of alternative testing approaches (6/13, 46%). This also was seen in searches for "Food intolerance Dublin," which had 57% (8/14) results endorsing non–evidence-based diagnostics.

Regional Differences in the Proportion of Nonevidence-Based Webpages Encountered

Comparison of the 2015 and 2019 results showed a marginal improvement in the availability of rational, scientific information regarding allergy over this timeframe. However, when the Irish search results from 2019 were evaluated against current search results for comparable UK and US populations, significant inadequacies are highlighted. Data were generated with identical search terms localizing to Southampton (United Kingdom) and Rochester (United States). Assessment of these results indicated that the vast majority of websites generated for each search pertained to websites promoting rational evidence-based approaches to assessment. A preponderance of academic and public hospital webpages was noted. Searches for "allergy Rochester New York," "allergy test Rochester New York," and "food allergy Rochester New York" resulted solely in rational testing approaches and scientifically robust information, without any non–evidence-based websites featuring.

Discussion

The Increasing Burden of Allergic Disease

The incidence of allergic conditions in both developed and developing countries has been increasing for over 50 years. One-third of people in the United Kingdom are estimated to suffer symptoms related to allergy at some stage during their lives [2]. The use of the internet has also risen steadily over recent generations, especially involving the investigation of medical conditions on the web. Over 70% of internet users in the United States have stated that they look on the web for health information, with over three-quarters of these queries beginning on a web-based search engine such as Google [14].

The increase in demand for web-based information creates both challenges and opportunities. Appropriate use of the internet enables health services to positively affect the lives of many people living with allergies. People who access health information on the internet have been shown to be more likely affected in their offline approach to health care [15], which in turn impacts on the provision of health care services. Figures on the extent of costs are difficult to obtain in Ireland; however, allergic disease accounts for 6% of general practitioner (GP) consultations in the National Health Service (NHS), 0.6% of hospital admissions, and 10% of GP prescribing budget. The cost (excluding hospital services) to the NHS is approximately £900 million (US $1.2 billion) per year [16]. Inadequately treated allergic disease also leads to a significant reduction in economic productivity with €55 to €151 billion (US $60 to $164 billion) lost per annum in the European Union, according to a 2014 study [17].

Complementary and Alternative Medicine Practitioners

When members of the Irish public search for guidance on the web regarding allergic disease, they are faced with a large volume of alternative testing approaches, often taking precedence over results for rational services. This is rightly a cause for concern. Previous studies in North America show that 91.5% of users will select a website from the first page and that the likelihood of a user clicking a result on the third page is 1.1% [18]. The ease of access to non–evidence-based information has the potential to promote costly practices and to increase the pressure on an already burdened health system through mismanagement of allergic diseases. There are countless medical conditions that attract input from non–evidence-based sources, but allergology has consistently shown itself to be an area of particular interest among CAM practitioners. A study of the websites of over 300 alternative health care providers demonstrated that 85% of naturopaths offer diagnosis, treatment, or efficacy for specifically treating allergic disease or sensitivity [19]. Similarly, a 2018 study investigating the educational quality of 300 food allergy YouTube videos showed that alternative medicine providers were the most common source of such content, with almost half of the videos depicting non–IgE-mediated reactions and frequently recommending controversial diagnostics [20]. In contrast to the word "allergy," which is frequently misappropriated by such websites, a recent study of worldwide internet search results for "anaphylaxis" showed that links to well-established, evidenced-based information were far more often seen [21].

Unmet Clinical Need in Allergic Disease

The internet has created a means of positively impacting the lives of many individuals affected by allergies while simultaneously decreasing the costs placed on the Irish health service. However, these potential benefits rely on the availability of accurate, evidence-based, and accessible resources that are provided by experts in allergology. There is a wide variation in the amount of allergologists working in different European countries, with a mean of 1.81 specialists per 100,000 inhabitants [22]. Figures for Ireland are not directly comparable as the country does not have a recognized training scheme for allergy. With a rate of 0.14 immunologists (pediatric and adult) per 100,000 citizens, who also have commitments to laboratory work, immunodeficiency, and autoimmunity, the relative time dedicated solely to allergy is assumed to be far below the European average [23]. Given that allergy is not currently a recognized medical specialty in Ireland, it should come as no surprise that tertiary allergy referral centers make up a relatively small proportion of localized web-based searches and that services run by GPs are often seen. Much has been previously

published on how best to alleviate the pressure on chronically underresourced tertiary allergy departments [24]. A common theme throughout these (predominantly UK-based) reports is improved support and education of primary care physicians in the field of allergy. Proposed interventions have included development of a network of GPs with special interest in allergy [25,26], a core allergy curriculum for all GPs [27], and better training of nurses, pharmacists, and dieticians to enable them to advise patients in the community [28].

Future Directions in Allergic Disease Management: Investment in Information Technology Resources

Although these proposals are commendable and steps should certainly be taken in their establishment, there can be no doubt that these are long-term investments and change will be slow to occur. There has also been growing interest in the use of information technology (IT)–based interventions, such as telemedicine assessments in adult allergy [29], app-based monitoring of allergic rhinitis [30], and a pilot program of email communication between allergists and nonspecialists for new referrals to allergy clinics [31], with further prospective studies required. Innovative and accessible approaches to delivering allergy services are required in the setting of an inadequately resourced system, where the lack of timely care for patients has undoubtedly contributed to people seeking out alternative practitioners. There is an undeniable need for increased support and funding for dedicated allergy services currently in operation in Ireland. However, to create a meaningful impact on patients' health-seeking behaviors, we need a parallel investment in IT services currently in use by the Irish health service.

In 2019, the information and communication technology capital allocation for the Irish health service was €85 million (US $92 million), making up approximately 0.5% of the total health care budget of over €16 billion (US $17 billion) [32]. The Health Service Executive (HSE) website was given a radical overhaul in 2013, which included development of a web-based database of over 600 health conditions and treatments. This database (entitled *Health A-Z*) contains useful information on the diagnosis and management of allergy and highlights the existence of alternative testing approaches, unambiguously describing them as unproven, unreliable, and best avoided by the public [33]. The content for this sizeable information resource was provided to the HSE completely free of charge from the NHS Choices website in the United Kingdom. Details on the volume of internet traffic to the HSE *Health A-Z* are not readily accessible to the public on the web. In our subsearch results for Dublin in 2019 (for "allergy test," "food allergy," and "food intolerance"), the HSE website never featured within the top 5 results. The question of cost efficacy and suboptimal internet traffic is certainly an issue that has been raised regarding the prototypic NHS Choices website in the past. A report published several years after its launch evaluated the NHS Choices website against the average website of a US company at the time (with a similar volume of monthly visitors) and found that the UK Department of Health was spending nearly four times more on site management, hosting costs, and development than their counterparts [34]. The downstream effect of this was the NHS entering a partnership with a digital marketing agency in 2012, which was specializing in analysis of consumer behaviors. This led to a significant improvement in their volume of web-based traffic [35]. The Irish health system, having saved huge costs in the development of their web-based health database, would undoubtedly benefit from external IT consultancy to improve their reach and strengthen their web-based presence. A public that is well informed and empowered in matters of their own health is a worthy long-term investment, and this applies to all areas of medicine, not solely allergology.

Conclusions

This study provides a snapshot of the information obtained when searching for information relating to allergy on the internet. There is great potential to provide accurate and evidence-based guidance to an increasing population, which would maximize the appropriate use of allergy services in Ireland. Unfortunately, this is not currently the case. Results for costly, non–evidence-based services predominate when searching on the web in Ireland, which differs relative to the United Kingdom and United States. There is a great need to improve the provision of allergy services in Ireland and to educate the Irish public on allergic disease, and investment in local internet resources is central to this endeavor.

Conflicts of Interest

None declared.

References

1. Central Statistics Office. 2018 Aug. Information Society Statistics-Households, 2018 URL: https://www.cso.ie/en/releasesandpublications/er/isshh/informationsocietystatistics-households2018/ [accessed 2019-07-30]
2. House of Lords. United Kingdom Parliament. 2006. Science and Technology Committee: Sixth Report of Session 2006-07 URL: https://publications.parliament.uk/pa/ld200607/ldselect/ldsctech/166/166ii.pdf [accessed 2019-07-30]
3. Irish Nutrition and Dietetics Institute. Food Allergies and Intolerances URL: https://www.indi.ie/diseases,-allergies-and-medical-conditions/food-allergy-and-intolerance/383-food-allergies-and-intolerances-factsheet.html [accessed 2019-07-30]
4. Irish Food Allergy Network. Food Allergy in Summary URL: http://ifan.ie/summary/ [accessed 2019-07-30]
5. Central Statistics Office. 2016 Nov. Irish Health Survey 2015 URL: https://www.cso.ie/en/releasesandpublications/ep/p-ihs/irishhealthsurvey2015/ct/ [accessed 2020-02-10]

6. Muraro A, Werfel T, Hoffmann-Sommergruber K, Roberts G, Beyer K, Bindslev-Jensen C, EAACI Food Allergy and Anaphylaxis Guidelines Group. EAACI food allergy and anaphylaxis guidelines: diagnosis and management of food allergy. Allergy 2014 Aug;69(8):1008-1025. [doi: 10.1111/all.12429] [Medline: 24909706]
7. Irish Food Allergy Network. IFAN & IAAI Issue Statement on Allergy and Food Intolerance Testing URL: https://www.indi.ie/health-professionals/690-ifan-iaai-issue-statement-on-allergy-and-food-intolerance-testing.html [accessed 2020-02-10]
8. Sampson HA, Aceves S, Bock SA, James J, Jones S, Lang D, Joint Task Force on Practice Parameters, Practice Parameter Workgroup, et al. Food allergy: a practice parameter update-2014. J Allergy Clin Immunol 2014 Nov;134(5):1016-25.e43. [doi: 10.1016/j.jaci.2014.05.013] [Medline: 25174862]
9. UK Advertising Standards Authority and Committee of Advertising Practice Executive. 2019 Feb 6. Health: Food Allergy Testing URL: https://www.asa.org.uk/advice-online/health-food-allergy-testing.html [accessed 2019-07-30]
10. Kemppainen LM, Kemppainen TT, Reippainen JA, Salmenniemi ST, Vuolanto PH. Use of complementary and alternative medicine in Europe: Health-related and sociodemographic determinants. Scand J Public Health 2018 Jun;46(4):448-455 [FREE Full text] [doi: 10.1177/1403494817733869] [Medline: 28975853]
11. Land MH, Wang J. Complementary and alternative medicine use among allergy practices: results of a nationwide survey of allergists. J Allergy Clin Immunol Pract 2018;6(1):95-8.e3. [doi: 10.1016/j.jaip.2017.01.017] [Medline: 28351787]
12. Kelso JM. Unproven diagnostic tests for adverse reactions to foods. J Allergy Clin Immunol Pract 2018;6(2):362-365. [doi: 10.1016/j.jaip.2017.08.021] [Medline: 29524991]
13. Google Trends. Explore: allergy test URL: https://trends.google.com/trends/explore?date=2009-01-01%202018-12-31&q=allergy%20test [accessed 2019-07-30]
14. Fox S, Duggan M. Pew Research Center. Washington, DC: Pew Internet and American Life Project; 2013 Jan 15. Health Online 2013 URL: https://www.pewinternet.org/2013/01/15/health-online-2013/ [accessed 2019-08-01]
15. Rainie L, Fox S. Pew Research Center. Washington, DC: Pew Internet and American Life Project; 2000 Nov 26. The Online Health Care Revolution URL: https://www.pewinternet.org/2000/11/26/the-online-health-care-revolution/ [accessed 2019-08-01]
16. Royal College of Physicians of London. Allergy: The Unmet Need. London: Royal College of Physicians of London; Jun 2003.
17. Zuberbier T, Lötvall J, Simoens S, Subramanian SV, Church MK. Economic burden of inadequate management of allergic diseases in the European Union: a GA(2) LEN review. Allergy 2014 Oct;69(10):1275-1279. [doi: 10.1111/all.12470] [Medline: 24965386]
18. Chitika Insights. 2013 Jun 7. The Value of Google Positioning URL: http://info.chitika.com/uploads/4/9/2/1/49215843/chitikainsights-valueofgoogleresultspositioning.pdf [accessed 2019-08-06]
19. Murdoch B, Carr S, Caulfield T. Selling falsehoods? A cross-sectional study of Canadian naturopathy, homeopathy, chiropractic and acupuncture clinic website claims relating to allergy and asthma. BMJ Open 2016 Dec 16;6(12):e014028 [FREE Full text] [doi: 10.1136/bmjopen-2016-014028] [Medline: 27986744]
20. Reddy K, Kearns M, Alvarez-Arango S, Carrillo-Martin I, Cuervo-Pardo N, Cuervo-Pardo L, et al. YouTube and food allergy: An appraisal of the educational quality of information. Pediatr Allergy Immunol 2018 Jun;29(4):410-416. [doi: 10.1111/pai.12885] [Medline: 29512839]
21. Kornafeld A, Gonzalez-Estrada A, Dimov V. 'Googling' anaphylaxis. Curr Opin Allergy Clin Immunol 2019 Oct;19(5):432-438. [doi: 10.1097/ACI.0000000000000575] [Medline: 31369414]
22. Fyhrquist N, Werfel T, Bilò MB, Mülleneisen N, van Wijk RG. The roadmap for the Allergology specialty and allergy care in Europe and adjacent countries. An EAACI position paper. Clin Transl Allergy 2019;9:3 [FREE Full text] [doi: 10.1186/s13601-019-0245-z] [Medline: 30697418]
23. Health Service Executive. Approved Consultant Establishment 30 September 2019 URL: https://www.hse.ie/eng/staff/leadership-education-development/met/consultantapplications/rep1/consultant-establishment-as-at-30-september-2019.pdf [accessed 2020-02-10]
24. Diwakar L, Cummins C, Lilford R, Roberts T. Systematic review of pathways for the delivery of allergy services. BMJ Open 2017 Feb 7;7(2):e012647 [FREE Full text] [doi: 10.1136/bmjopen-2016-012647] [Medline: 28174222]
25. Levy ML, Walker S, Woods A, Sheikh A. Service evaluation of a UK primary care-based allergy clinic: quality improvement report. Prim Care Respir J 2009 Dec;18(4):313-319 [FREE Full text] [doi: 10.4104/pcrj.2009.00042] [Medline: 19588053]
26. El-Shanawany IR, Wade C, Holloway JA. The impact of a General Practitioner-led community paediatric allergy clinic: a service evaluation. Clin Exp Allergy 2019 May;49(5):690-700. [doi: 10.1111/cea.13375] [Medline: 30791174]
27. Royal College of Physicians. Allergy Services: Still Not Meeting the Unmet Need. London: Royal College of Physicians; Jun 2010.
28. Ryan D, Levy M, Morris A, Sheikh A, Walker S. Management of allergic problems in primary care: time for a rethink? Prim Care Respir J 2005 Aug;14(4):195-203 [FREE Full text] [doi: 10.1016/j.pcrj.2005.01.003] [Medline: 16701725]
29. Krishna MT, Knibb RC, Huissoon AP. Is there a role for telemedicine in adult allergy services? Clin Exp Allergy 2016 May;46(5):668-677. [doi: 10.1111/cea.12701] [Medline: 26742680]

30. Bousquet J, Schunemann HJ, Fonseca J, Samolinski B, Bachert C, Canonica GW, et al. MACVIA-ARIA Sentinel NetworK for allergic rhinitis (MASK-rhinitis): the new generation guideline implementation. Allergy 2015 Nov;70(11):1372-1392. [doi: 10.1111/all.12686] [Medline: 26148220]
31. Conlon NP, Abramovitch A, Murray G, O'Hanrahan A, Wallace D, Holohan K, et al. Allergy in Irish adults: a survey of referrals and outcomes at a major centre. Ir J Med Sci 2015 Jun;184(2):349-352. [doi: 10.1007/s11845-014-1117-0] [Medline: 24744258]
32. Health Service Executive. National Service Plan 2019 URL: https://www.hse.ie/eng/services/publications/serviceplans/national-service-plan-2019.pdf [accessed 2019-09-20]
33. Health Service Executive. Allergy, Food URL: https://www.hse.ie/eng/health/az/a/allergy,-food/causes-of-a-food-allergy.html [accessed 2019-09-20]
34. UKFast. NHS Choices Website Costs More Than £27m A Year URL: https://www.ukfast.co.uk/hosting-news/more-than-pound27m-spent-annually-on-nhs-choices-website.html [accessed 2019-09-20]
35. Hitwise. NHS Choices Gets Relevant, Agile, and Accessible with Hitwise URL: https://www.hitwise.com/en/2018/12/12/case-study-nhs-choices/ [accessed 2019-09-20]

Abbreviations

CAM: complementary and alternative medicine
GP: general practitioner
HSE: Health Service Executive
IT: information technology
NHS: National Health Service

Edited by G Eysenbach; submitted 22.10.19; peer-reviewed by A Allam, D Ryan, D Dalan; comments to author 17.12.19; revised version received 31.12.19; accepted 26.01.20; published 13.05.20.

Please cite as:
King C, Judge C, Byrne A, Conlon N
Googling Allergy in Ireland: Content Analysis
J Med Internet Res 2020;22(5):e16763
URL: https://www.jmir.org/2020/5/e16763
doi:10.2196/16763
PMID:32401220

Original Paper

Classification and Prediction of Violence Against Chinese Medical Staff on the Sina Microblog Based on a Self-Organizing Map: Quantitative Study

Guimin Duan[1], PhD; Xin Liao[2], MA; Weiping Yu[3], PhD; Guihua Li[4], PhD

[1]School of Management, Chengdu University of Traditional Chinese Medicine, Chengdu, Sichuan, China

[2]School of Public Affairs and Law, Southwest Jiaotong University, Chengdu, Sichuan, China

[3]School of Business, Sichuan University, Chengdu, Sichuan, China

[4]School of Public Administration, Sichuan University, Chengdu, Sichuan, China

Corresponding Author:
Xin Liao, MA
School of Public Affairs and Law
Southwest Jiaotong University
No 111, First Section, North of Second Ring Road
Chengdu, Sichuan, 610031
China
Phone: 86 17761305794
Email: liao_shane@163.com

Abstract

Background: For the last decade, doctor-patient contradiction in China has remained prominent, and workplace violence toward medical staff still occurs frequently. However, little is known about the types and laws of propagation of violence against medical staff online.

Objective: By using a self-organizing map (SOM), we aimed to explore the microblog propagation law for violent incidents in China that involve medical staff, to classify the types of incidents and provide a basis for rapidly and accurately predicting trends in public opinion and developing corresponding measures to improve the relationship between doctors and patients.

Methods: For this study, we selected 60 cases of violent incidents in China involving medical staff that led to heated discussions on the Sina microblog from 2011 to 2018, searched the web data of the microblog using crawler software, recorded the amount of new tweets every 2 hours, and used the SOM neural network to cluster the number of tweets. Polynomial and exponential functions in MATLAB software were applied to predict and analyze the data.

Results: Trends in the propagation of online public opinion regarding the violent incidents were categorized into 8 types: bluff, waterfall, zigzag, steep, abrupt, wave, steep slope, and long slope. The communications exhibited different characteristics. The prediction effect of 4 types of incidents (ie, bluff, waterfall, zigzag, and steep slope) was good and accorded with actual spreading trends.

Conclusions: Our study found that the more serious the consequences of a violent incident, such as a serious injury or death, the more attention it drew on the microblog, the faster was its propagation speed, and the longer was its duration. In these cases, the propagation types were mostly steep slope, long slope, and zigzag. In addition, the more serious the consequences of a violent incident, the higher popularity it exhibited on the microblog. The popularity within a week was significantly higher for acts resulting from patients' dissatisfaction with treatments than for acts resulting from nontherapeutic incidents.

(J Med Internet Res 2020;22(5):e13294) doi:10.2196/13294

KEYWORDS

workplace violence; medical staff; social media

Introduction

Background

At the beginning of 2009, China started a new round of medical and health system reformation. Its guiding ideology was to continuously improve the health level of the whole population and promote social harmony, which has improved, to a great extent, the doctor-patient relationship. However, contradictions between doctors and patients are still prominent, and workplace violence toward medical staff occur frequently. The Chinese Hospital Association reported that the proportion of hospitals experiencing workplace violence increased by 90% from 2008 to 2012 [1]. In 2017 itself, there were as many as 54 violent incidents against Chinese medical staff mentioned on the Sina microblog (a leading microblog website and one of the largest social media platforms in China), which attracted public attention and sparked heated debate. For instance, on December 8, 2017, in Tongshan County People's Hospital, a patient's mother hurt a nurse because she was not satisfied with the treatment effect, causing slight bodily injury (signs of abortion) to the medical staff. When the incident was publicized on the Sina microblog, the number of tweets reached 1737 within a week. The frequent outbreak of violent incidents in hospitals and the wide coverage of the incidents by social media have made the doctor-patient relationship even more fraught, which not only reduces the professional security of medical staff [2] but also hinders their professional performance, resulting in a negative impact on patient service and overall health [3-8]. Moreover, violent incidents have a negative impact on the psychological welfare of health care workers who do not want their children to be engaged in health care [9-11].

Workplace violence in the health sector is a worldwide concern, with health care workers being at high risk of being victims [12]. Following the World Health Organization's definition, workplace violence takes 2 main forms: physical and nonphysical violence. Physical violence includes hitting, slapping, kicking, pushing, choking, grabbing, sexual assault, and other forms of physical contact intended to injure or harm. In contrast, nonphysical violence includes threats, sexual harassment, bullying, and verbal abuse and may be perpetrated by various types of people [13]. There are 2 major sources of workplace violence: coworker initiated and public initiated [14]. In the Chinese health sector, workplace violence is mainly caused by patients and their relatives or friends. Scholars have studied the causes of violent incidents from various perspectives. On the basis of the perspective of social roles, Robinson [15] pointed out that due to the differences in the roles of doctors and patients, their viewpoints and interests differ, which leads to cognitive and motivational bias in the attribution process regarding medical information. From the point of view of social psychology, Feng et al [4] analyzed the mechanisms for the inducement of violent incidents related to doctor-patient conflicts, including psychological frustration, anger, social learning, and so on. From the perspective of patient experience, Hu et al [5] proposed that the experience of a poor physical environment, inadequate information communication, large medical expenses, and the perception and experience of treatment outcomes were direct causes of doctor-patient contradiction. From the perspective of crisis management, Li [6] posited that it was caused by conflicts of interest, the inadequate communication of information, negative slants in public opinion, and so on. Duan et al [16] used content analysis to classify violent incidents and associated them with 6 factors: diagnosis and treatment effects, doctor-patient communication, response speed, medical expenses, privacy protection, and patients' or their families' problems.

Objectives

With the development of social media, the microblog has become a key vehicle of public opinion by virtue of its extensive and real-time capabilities for mass communication. Despite the widespread coverage of violent incidents toward medical staff on social media, little empirical research has examined their types and laws of propagation. Although scholars have studied public opinion about violent incidents, they have focused mainly on influencing factors, the dynamic mechanisms of public opinion, and the characteristics of the incidents [17,18]. Some scholars have conducted research on the types and characteristics of microblog communication in light of corporate crisis events [19]. However, due to the particularity of medical services and the diversity of incentives for committing violent acts against medical staff, existing research results cannot effectively guide hospitals and health management departments in managing public opinion about the incidents. In relation to this issue, we took violent incidents against Chinese medical staff as the research object and applied two methods—the self-organizing map (SOM) neural network and function fitting—to conduct a quantitative study. The purpose of this study was to explore the microblog propagation law for violent incidents in China that involve medical staff, to classify the types of incidents, and to provide a basis for rapidly and accurately predicting trends in public opinion and developing corresponding measures to improve the relationship between doctors and patients. We selected 60 cases of violent incidents that led to heated discussions on the Sina microblog from 2011 to 2018 and used crawler software to search the microblog's web data. Starting from the first post about the incidents, every 2 hours, we recorded the number of additional tweets. We used the SOM to demarcate changes in the number of tweets and used the exponential and polynomial functions in MATLAB R2017a (The MathWorks) to predict and analyze the data to achieve better prediction. For hospitals and health management departments, understanding the communication types and the characteristics of the violent incidents being discussed on microblogs such as the Sina microblog can be helpful to institutions for scientifically predicting the propagation rules of incidents, for effectively managing and guiding public opinion relating to doctors and patients, and for improving the doctor-patient relationship.

Methods

Data Collection and Processing

We used "doctor/nurse was hit," "doctor/nurse was slapped," "doctor/nurse was killed," "doctor/nurse was attacked," "doctor/nurse was hurt," and so on as keywords when searching for violent incidents from January 1, 2011, to July 31, 2018,

using the Baidu search engine, and we searched for complete data on the development and outcomes of the target incidents. In total, we collected 403 violent incidents involving Chinese medical staff. According to Duan's classification criteria, we used the content analysis method to analyze the inducements for 60 violent incidents and grouped them into 6 categories: treatment effect, doctor-patient communication, response speed, medical expenses, privacy protection, and patients' or families' own problems [16]. Relying on the crawler software platform of GooSeeker (Shenzhen Tianju Information Technology Ltd), we defined the fetching rules, searched related tweets by keyword, and then collected the website data from the Sina microblog to record the total number of tweets in a week.

As the propagation characteristics of violent incidents against Chinese medical staff on the Sina microblog are not obvious and the degree of influence is relatively low, we considered the total number of tweets of violent incidents on the Sina microblog within a week as the standard and selected 60 typical incidents that drew heated responses and discussion among netizens from 2011 to 2018 as the study samples. The samples covered the 5 direct causes of violent incidents involving medical staff and were representative to some extent. In accordance with the "criteria for human injury evaluation" issued by the Judicial Department in China, the severity of the consequences of violent incidents was classified as "slight bodily injury," "minor injury," "serious injury," "death," or "serious influence on hospital order" (as listed in Multimedia Appendix 1).

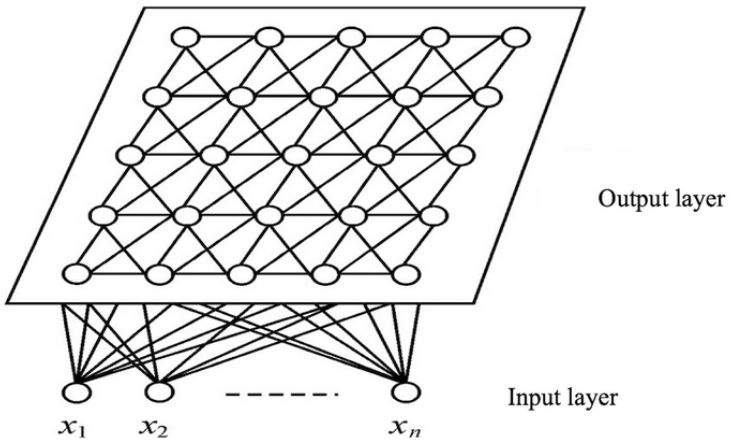
We mainly studied the dynamics of public opinion in the first week after the violent incidents. Therefore, we started to record tweets about the first violent incident reports and then recorded newly added Sina microblog material every 2 hours. We searched 168 hours of data within a week and collected 84 datasets in total. There were differences in the number of tweets for different events, so the data were standardized.

The formula was as follows: adjusted standard value=100×(original value−min)/(max−min). Min and max are the minimum and maximum values for each of the 84 datasets about the incidents.

Network Sample Design

The SOM is an excellent tool in the exploratory phase of data mining [20]. SOM networks learn to cluster groups of similar input patterns from a high-dimensional input space in a nonlinear fashion to a low-dimensional (most commonly two-dimensional) discrete lattice of neurons in an output layer [21], which was widely applied to classification and prediction [22-27]. The neurons in the competition layer can be one-dimensional, two-dimensional, or multidimensional (as shown in Figure 1 [28]). SOMs are reported to be robust and accurate with noisy data [29].

Figure 1. A topology diagram of the self-organizing map network.



The SOM method works as follows [30]. Initially, one has to choose the topology of the map. All the nodes are linked to the input nodes by weighted edges. The weights are first set at random and then iteratively adjusted. Each iteration involves randomly selecting an object x and moving the closest node (and its neighborhood) in the direction of x. The closest node is obtained by measuring the Euclidean distance or the dot product between the object x and the weights of all nodes in the map. The neighborhood to be adjusted is defined by a neighborhood function, which decreases over time.

We used the MATLAB programming language to construct the SOM neural network model and employed functions provided by a neural network toolbox to implement the whole learning process, such as training, emulating functions, and so on. According to the data after standardized processing, the input mode of the network was determined as

where k=1,2,3,…,60 (n=84), that is, there were 60 sets of samples, each of which contained 84 elements.

Results

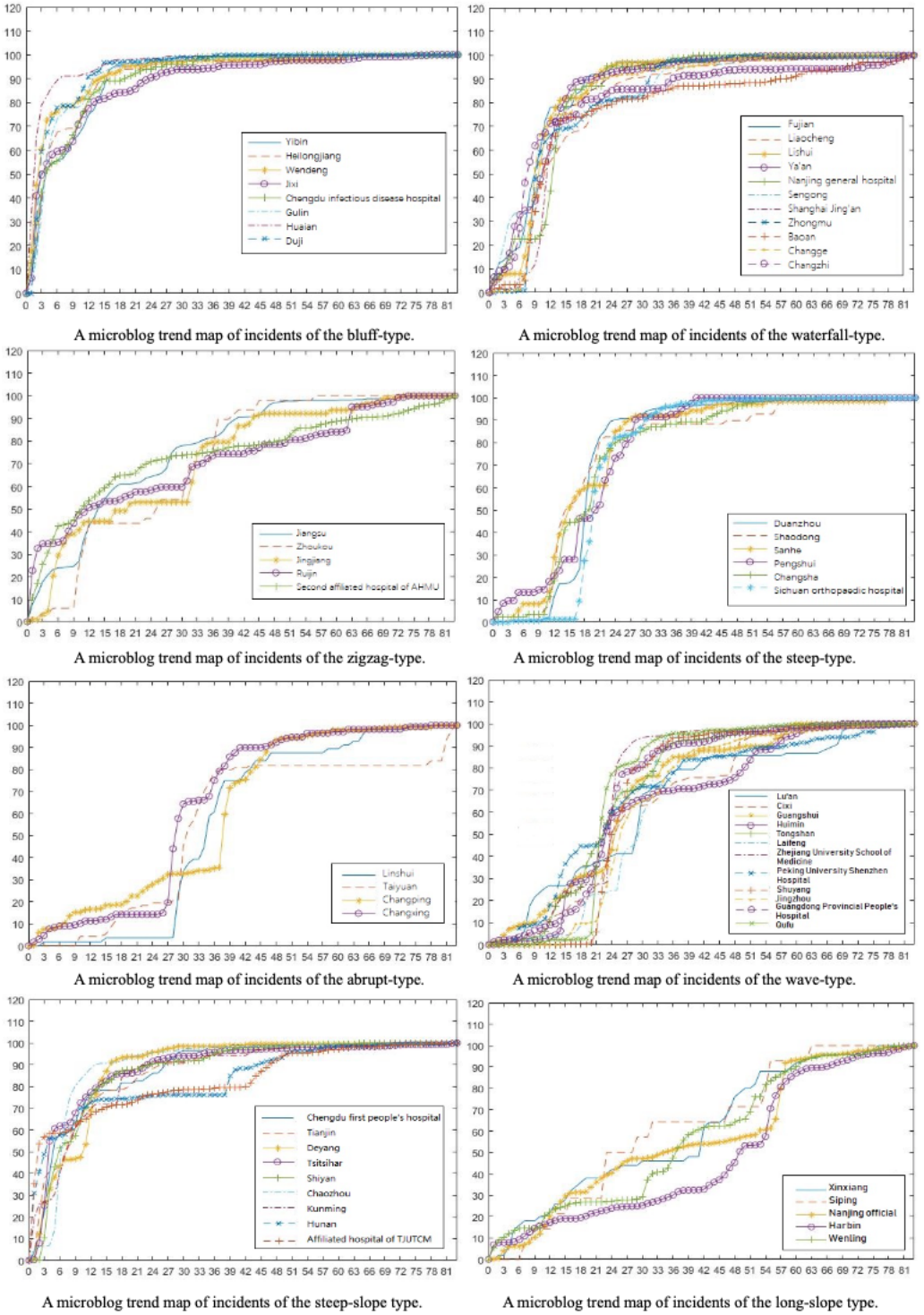
Clustering Results

With the SOM neural network toolkit of MATLAB R2017a, 60 groups of data, containing 84 elements each, were clustered by standardized processing into 8 types. On the basis of the shapes of the trend lines in the graphs, the types of clustering results were named as follows: *bluff*, *waterfall*, *zigzag*, *steep*, *abrupt*, *wave*, *steep slope*, and *long slope* (shown in Figure 2).

By the application of the SOM neural network algorithm, 60 medical incidents against Chinese medical staff were divided

into 8 types, based on the number of dynamic changes in the Sina microblog tweets. In Figure 2, it can be seen that these 8 types of curves generally showed significant differences, albeit similar curves represent similar trends. The types were all named according to the general shapes of the curves. It can be seen from the figures that the microblog transmission curves of each type had distinguishing characteristics (shown in Table 1). The clustering results of the SOM neural network model were good.

Figure 2. A microblog trend map of incidents of the types.



AHMU: Anhui Medical University

TJUTCM: Tianjin University of Traditional Chinese Medicine

Table 1. Clustering results and curve characteristics.

Category	Hot violent incidents with medical staff	Curve characteristics
Bluff type	Yibin Second People's Hospital, December 17, 2017; Heilongjiang Provincial People's Hospital, August 15, 2017; Shandong Wendeng Osteopathic Hospital, March 21, 2017; Jixi People's Hospital, July 11, 2017; Chengdu Infectious Disease Hospital, July 21, 2017; Gulin County People's Hospital, January 14, 2015; Huai'an First People's Hospital, February 10, 2018; Du Ji people's Hospital, March 21, 2016	In the first 12 hours or so, there was a peak period of tweet postings. After which, in the 13th hour, the growth rate suddenly declined and then lasted for 6 hours. After that, in the 19th hour, the number of tweets sped up, and the increase lasted for 6 hours or so. Then, it continued to increase at a slower pace. After the 96th hour, the growth rate was almost zero.
Waterfall type	Fujian Provincial Cancer Hospital, February 7, 2017; Liaocheng People's Hospital, December 4, 2017; Hong Lan Town Health Centre of Lishui District, April 23, 2017; Ya'an People's Hospital, January 5, 2017; Nanjing General Hospital of Nanjing Command, January 17, 2017; Shanxi Sengong Hospital, February 1, 2017; Jing'an District Central Hospital, March 4, 2016; Yanminghu Central Health Center of Zhongmu County, June 11, 2015; Baoan District people's Hospital, September 9, 2013; Changge People's Hospital, November 1, 2015; Heping Hospital affiliated to Changzhi Medical College, November 22, 2016	The early growth rate was slow. At around 12 hours, the growth rate of the microblog increased sharply and then gradually slowed down.
Zigzag type	Jiangsu Provincial People's Hospital, February 16, 2017; Zhoukou Central Hospital, July 12, 2017; Orthopaedic Clinic in Jingjiang City, April 22, 2017; Ruijin Hospital Affiliated to Medical College of Shanghai Jiaotong University, June 27, 2015; Second Affiliated Hospital of Anhui Medical University, November 13, 2012	Growth was rapid in the early stage, and then the growth trend became slow, or even close to 0, followed by a growth phase featuring alternating speeds.
Steep type	Women and Children Health Hospital of Duanzhou District, August 6, 2017; Shaodong County People's Hospital, July 17, 2017; Sanhe People's Hospital, March 21, 2017; Pengshui County People's Hospital, March 24, 2017; Jiangbei Town Center Hospital of Changsha County, May 17, 2016; Sichuan Orthopaedic Hospital, October 10, 2016	In the early stages, the blog was in an incubation period. Then, there was a peak period of tweets posted between the 24th and 48th hours, after which the growth gradually slowed down. After the 114th hour, the growth rate was almost 0.
Abrupt type	Traditional Chinese Medicine Hospital of Linshui County, August 5, 2017; Taiyuan Central Hospital, August 5, 2017; Dongguan Changping Hospital, April 17, 2016; Changxing County People's Hospital, October 4, 2016	The growth rate was slow in the first 30 hours or so, and then, it increased slightly over the next 2 hours or so. Before the 54th hour, the growth rate was almost flat. Then, a peak period of posting occurred between 54 and 72 hours, after which the growth rate gradually slowed down.
Wave type	Lu'an People's Hospital, September 1, 2017; Cixi people's Hospital, June 4, 2017; Guangshui First People's Hospital, April 6, 2017; Huimin People's Hospital, June 15, 2017; Tongshan County People's Hospital, December 8, 2017; Lai Feng County Central Hospital, November 2, 2017; The Second Affiliated Hospital of Zhejiang University School of Medicine, February 20, 2014; Peking University Shenzhen Hospital, February 25, 2014; Shuyang County Nanguan Hospital, April 19, 2014; Jingzhou First People's Hospital, May 5, 2012; Guangdong Provincial People's Hospital, May 5, 2016; Qufu People's Hospital, May 16, 2016	The number of tweets increased gradually over the first 42 hours and then increased suddenly. Then, the growth rate showed a wave-shaped oscillation within 43 to 149 hours. After the 150th hour, the growth rate was almost 0.
Steep slope type	First People's Hospital of Chengdu, March 23, 2017; Tianjin Third Central Hospital, June 29, 2017; Deyang People's Hospital, January 3, 2016; Beigang Hospital, February 17, 2014; Shiyan People's Hospital, February 21, 2015; Chaozhou Central Hospital, March 5, 2014; First Affiliated Hospital of Kunming Medical University, June 7, 2015; Hunan Academy of Traditional Chinese Medicine Affiliated Hospital, September 23, 2013; First Teaching Hospital of Tianjin University of Traditional Chinese Medicine, November 29, 2012	The number of tweets increased sharply over the first 6 hours or so. Then, there was a short pause, followed by a gradual and steady increase from the 84th hour to the 101st hour. After the 102nd hour, the growth rate was almost 0.
Long slope type	Xinxiang Second People's Hospital, November 7, 2017; Siping Central Hospital, October 15, 2017; Nanjing Stomatological Hospital, February 25, 2014; First Affiliated Hospital of Harbin Medical University, March 23, 2012; Wenling First People's Hospital, October 25, 2013	The growth rate of the tweets was stable and lasted for a long time. After 120 hours, it gradually increased.

Correlation Analysis Between the Types of Propagation on the Microblog and the Characteristics of the Medical Incidents

The characteristics of the medical incidents against Chinese staff analyzed in this study were the severity of the consequences and the inducements. To explore the relationships between the types of propagation on the microblog and the severity of the consequences of the medical incidents, we conducted a one-way analysis of variance (ANOVA; single factor) considering the types of propagation on the microblog as the independent variables and severity of the consequence of the medical

incidents as the dependent variable. After administering the Shapiro-Wilk test, the dependent variable was normally distributed. The results show that there were significant differences in the severity of the consequences of incidents associated with the different types of propagation (P=.03). The severity of the consequences of medical incidents of the long slope type (mean 4.00) was significantly higher than those of the bluff type (mean 2.63), waterfall type (mean 2.73), steep type (mean 2.17), abrupt type (mean 2.5), and wave type (mean 2.58). The severity of the consequences of medical incidents of the steep slope type (mean 3.75) was significantly higher than those of the bluff type (mean 2.63), steep type (mean 2.17), and wave type (mean 2.58). The severity of the consequences of medical incidents of the zigzag type (mean 3.80) was significantly higher than those of the steep type (mean 2.17) and wave type (mean 2.58), as shown in Multimedia Appendix 2.

The abovementioned analysis shows that the severity of the consequences of the medical incidents varies with the type of propagation. The severity of incidents of the long slope type, steep slope type, and zigzag type was higher than that of the other types, and most of the incidents in these categories involved a serious injury or death. Therefore, we classified these incidents as a high-severity group and the other incidents as a low-severity group. We used a one-way ANOVA, with these 2 groups, that characterized the consequences of the medical incidents as the independent variable and popularity within a week of the incidents as the dependent variable. Popularity was rated on a 6-point scale according to the number of tweets within a week. We coded 6 for a number of tweets above 5000, 5 for a number from 2000 to 5000, 4 for 1000 to 2000, 3 for 500 to 1000, 2 for 100 to 500, and 1 for 0 to 100. The higher the point score, the higher the popularity. Therefore, the dependent variable is continuous. After administering the Shapiro-Wilk test, the dependent variable was normally distributed. The results show that the popularity of high-severity incidents was significantly higher than that of low-severity incidents (high: mean 3.44; low: mean 2.05; P<.001). We adopted a one-way ANOVA, with the types of inducement for the incidents as independent variables and popularity within a week as the dependent variable. We found that violent acts resulting from patients' dissatisfaction with treatment effects garnered popularity within a week that was significantly higher than the popularity of incidents caused by nontherapeutic effects (treatment effect: mean 3.05; nontherapeutic effect: mean 2.15; P=.02).

We constructed a cross-tabulation of the inducements of the medical incidents and the types of propagation on the microblog. The chi-square test results show that there was no correlation between them.

Forecasting Trends in Network Public Opinion Regarding Violent Incidents

Nowadays, microblogs are becoming important carriers of public opinion on the internet about violent incidents against medical staff, and the traditional methods that medical institutions and health management departments use to control public opinion are severely weakened. The dynamic monitoring of trends and changes in microblog discussions can help hospital and health management departments to take timely countermeasures after violent incidents are publicized. At the same time, relevant departments could make use of the data to predict the propagation of microblog discussions of violent incidents, effectively manage and guide public opinion about doctors and patients, and improve the relationship between doctors and patients. On the basis of the study on stock price prediction by Afolabi and Olude [31], we considered 4 types of propagation, including the bluff, waterfall, zigzag, and steep slope types, for the prediction analysis in this study.

Forecasting Trends in the Spread of Network Public Opinion: Bluff-Type Propagation

The first cluster, consisting of incidents with the bluff-type profile, was selected for the sample. We took the average data of the 8 incidents in the Sina microblog and used the cftool toolkit in MATLAB R2017a to do the exponential function fitting for prediction. We recorded newly added Sina microblog material every 2 hours and searched 168 hours of data within a week, collecting 84 datasets in total. The first 70 of the 84 datasets were selected for fitting. We determined the fitting model and used the last 14 datasets as the prediction results to validate the data. As shown in Figure 3, *fit 1* was the fitting curve, and y vs x was the original data point.

The fitting model for the exponential function was

$$f(x_1) = 93.39 \times \exp(0.001092 \times x_1) - 112.8 \times \exp(-0.204 \times x_1)$$

where x_1 was the data observation point (we took observations every 2 hours) and $f(x_1)$ was the number of related tweets after standardized processing under the observation point.

In the fitting of the model, R^2 (the coefficient of determination) was 0.985, the sum of squares due to error (SSE) was 377.7, adjusted R^2 was 0.9846, and the root-mean-square error (RMSE) was 2.392.

As shown in Table 2, the predicted data values were almost consistent with the actual values, with a maximum relative error of 1.9055% and a minimum relative error of 0.77%. The trend of the data in the forecast section (positions 71 to 84) completely coincided with the actual trend, which shows that the prediction effect of the model for the bluff-type incidents was good.

Figure 3. A data-fitting diagram of the bluff-type incidents.

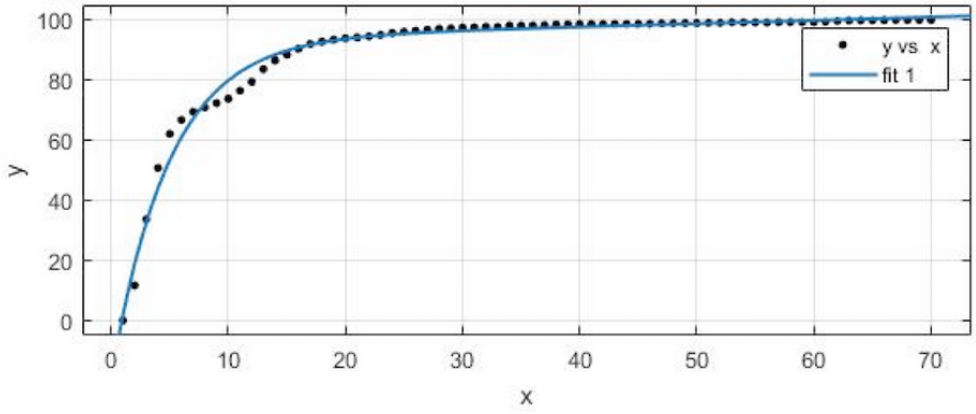


Table 2. A comparison of the fitting value with the actual value of the bluff-type incidents.

Predicted position	Fitting value	Actual value	Relative error (%)
71	100.5967	99.8280	0.7700
72	100.6968	99.8590	0.8390
73	100.7970	99.8918	0.9062
74	100.8973	99.9188	0.9793
75	100.9976	99.9188	1.0796
76	101.0981	99.9729	1.1255
77	101.1987	99.9729	1.2261
78	101.2994	99.9729	1.3268
79	101.4001	100.0000	1.4001
80	101.5010	100.0000	1.5010
81	101.6020	100.0000	1.6020
82	101.7031	100.0000	1.7031
83	101.8042	100.0000	1.8042
84	101.9055	100.0000	1.9055

Forecasting Trends in the Spread of Network Public Opinion: Waterfall-Type Propagation

The second cluster, consisting of waterfall-type incidents, was selected for the sample. We took the average data of the 11 incidents in the Sina microblog for the forecast and made use of the cftool toolkit in MATLAB R2017a to do the polynomial function fitting for prediction. As shown in Figure 4, *fit 2* was the fitting curve.

The fitting model of the polynomial function was

$$f(x_2) = (-1.156e\text{-}05)x_2^4 + 0.002763x_2^3 - 0.2373x_2^2 + 8.783x_2 - 22.33$$

where x_2 was the data observation point (we made an observation every 2 hours) and $f(x_2)$ was the value of the number of related tweets after standardized processing under the observation point.

In the fitting of the model, R^2 was 0.977, the SSE was 1356, adjusted R^2 was 0.9756, and RMSE was 4.568.

As shown in Table 3, the predicted data values were almost consistent with the actual values, with a maximum relative error of 3.06% and a minimum relative error of 1.3401%. The trend of the data in the forecast section (positions 71 to 84) completely coincided with the actual trend, which shows that the prediction effect of the model for the waterfall-type incidents was good.

Figure 4. A data-fitting diagram of the waterfall-type incidents.

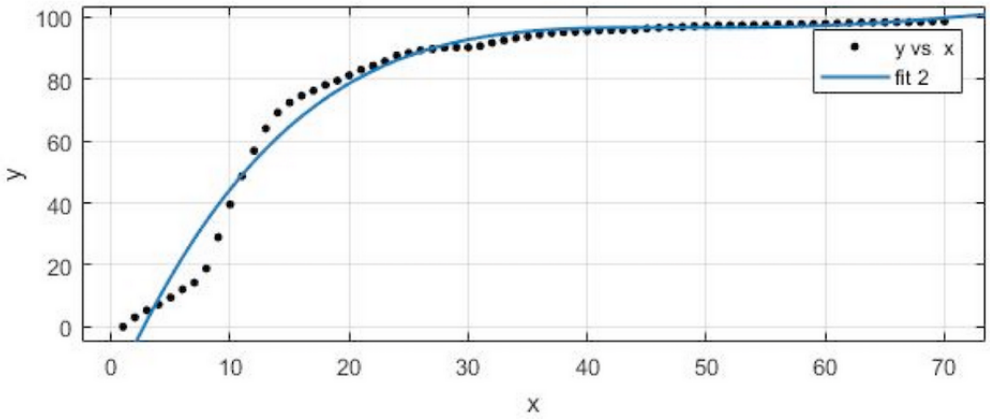


Table 3. A comparison of the fitting value with the actual value of the waterfall-type incidents.

Predicted position	Fitting value	Actual value	Relative error (%)
71	100.1828	98.8402	1.3401
72	100.5052	98.9306	1.5666
73	100.8276	99.0648	1.7483
74	101.1463	99.1315	1.9919
75	101.4575	99.2729	2.1532
76	101.7571	99.2850	2.4294
77	102.0408	99.2850	2.7007
78	102.3039	99.3088	2.9277
79	102.5415	99.4037	3.0600
80	102.7484	99.9145	2.7581
81	102.9191	99.9609	2.8743
82	103.0478	99.9609	2.9956
83	103.1286	100.0000	3.0337
84	103.1551	100.0000	3.0586

Forecasting Trends in the Spread of Network Public Opinion: Zigzag-Type Propagation

The third cluster, consisting of zigzag-type incidents, was also selected for the sample. For the forecast, we took the average data of the 5 incidents discussed in the Sina microblog and made use of the cftool toolkit in MATLAB R2017a to do the polynomial function fitting for prediction. As shown in Figure 5, *fit 3* was the fitting curve.

The fitting model of the polynomial function was

$$f(x_3) = 0.0001973x_3^3 - 0.04018x_3^2 + 3.166x_3 + 5.273$$

where x_3 was the data observation point (we made an observation every 2 hours) and $f(x_3)$ was the value of the number of related tweets after standardized processing under the observation point.

In the fitting of the model, R^2 was 0.9872, the SSE was 584.7, adjusted R^2 was 0.9866, and RMSE was 2.977.

As shown in Table 4, the predicted data values were almost consistent with the actual values, with a maximum relative error of 4.441% and a minimum relative error of 0.8448%. The trend of the data in the forecast section (positions 71 to 84) completely coincided with the actual trend, which shows that the prediction effect of the model for the zigzag-type incidents was good.

Figure 5. A data-fitting diagram of the zigzag-type incidents.

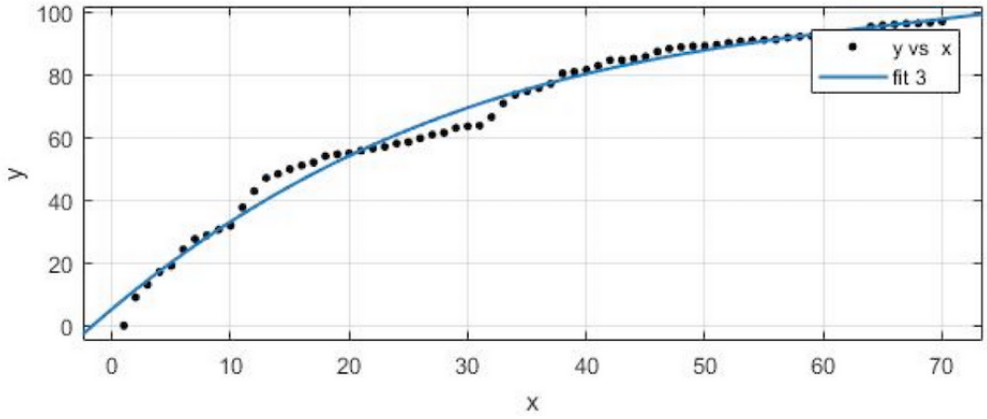


Table 4. Comparison of the fitting value with the actual value of the zigzag-type incidents.

Predicted position	Fitting value	Actual value	Relative error (%)
71	98.1275	97.2985	0.8448
72	98.5737	97.5108	1.0783
73	99.0248	98.0012	1.0336
74	99.4820	98.2928	1.1954
75	99.9464	98.6039	1.3432
76	100.4193	98.8007	1.6118
77	100.9017	99.0006	1.8841
78	101.3950	99.1359	2.2280
79	101.9002	99.2251	2.6252
80	102.4186	99.2435	3.1001
81	102.9513	99.3665	3.4820
82	103.4996	99.6064	3.7616
83	104.0646	99.7724	4.1245
84	104.6474	100.0000	4.4410

Forecasting Trends in the Spread of Network Public Opinion: Steep Slope–Type Propagation

The seventh cluster, consisting of steep slope–type incidents, was selected. For the forecast, we took the average data of the 9 incidents in the Sina microblog and made use of the cftool toolkit in MATLAB R2017a to do the exponential function fitting for prediction. As shown in Figure 6, *fit 4* was the fitting curve.

The fitting model of the exponential function was

$$f(x_4) = 85.79 \times \exp(0.002299 \times x_4) - 100 \times \exp(-0.1424 \times x_4)$$

where x_4 was the data observation point (we made an observation every 2 hours) and $f(x_4)$ was the value of the number of related tweets after standardized processing under the observation point.

In the fitting of the model, R^2 was 0.9987, the SSE was 44.51, adjusted R^2 was 0.9986, and RMSE was 0.8212.

As shown in Table 5, the predicted data values were almost consistent with the actual values, with a maximum relative error of 3.9058% and a minimum relative error of 1.5555%. The trend of the data in the forecast section (positions 71 to 84) completely coincided with the actual trend, which shows that the prediction effect of the model for the steep slope–type incidents was good.

Figure 6. A data-fitting diagram of the steep slope–type incidents.

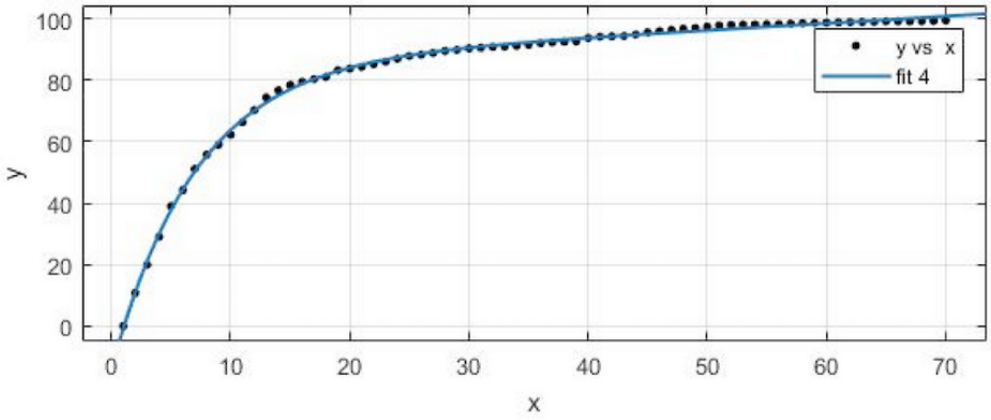


Table 5. A comparison of the fitting value with the actual value for the steep slope–type incidents.

Predicted position	Fitting value	Actual value	Relative error (%)
71	100.9970	99.4260341	1.5555
72	101.2300	99.4703353	1.7383
73	101.4635	99.5033538	1.9319
74	101.6975	99.5719260	2.0901
75	101.9319	99.5832087	2.3042
76	102.1668	99.6310762	2.4819
77	102.4022	99.6367044	2.7006
78	102.6382	99.6995552	2.8631
79	102.8746	99.7121893	3.0740
80	103.1116	99.7550284	3.2553
81	103.3490	99.8030223	3.4311
82	103.5870	99.8845621	3.5742
83	103.8256	99.9749575	3.7088
84	104.0646	100.0000000	3.9058

Discussion

Principal Findings

The theories and methods of public opinion communication are relatively mature, but applications in the field of medical public opinion are relatively few. We believe that this study is the first to apply the theories and methods of public opinion communication to research in medical public opinion and to systematically explore the types of incidents against medical staff and the patterns in which discussions about them propagate on microblogs. Tracking discussions of violent acts against medical staff in China that took place on the internet from 2011 to 2018, we selected 60 cases for our sample that prompted intense discussion on the Sina microblog, used crawler software to collect the data from the web, applied the SOM neural network model to cluster the tweets, mapped changes in the discussion trends over time, and used the polynomial and exponential function methods to analyze and predict communication patterns of public opinion. The results showed that the fit was good.

Our study found that the propagation of web-based public opinion relating to the violent incidents was divided into 8 types: bluff, waterfall, zigzag, steep, abrupt, wave, steep slope, and long slope. Over the first 12 hours or so, the number of tweets in the bluff type of discussion peaked, after which the growth rate declined suddenly in the 13th hour and continued to decline for 6 hours. In the 19th hour, there was an increase in the number of tweets that lasted for 6 hours or so, and then, the growth rate gradually decreased. After the 96th hour, the growth rate was almost 0. For the waterfall-type propagation, the early growth rate was slow. At about the 12th hour, the growth rate increased sharply, then gradually decreased. For the zigzag type of propagation, growth was rapid in the early stage, and then, it decreased and sometimes approached 0, going through a process of alternating slow and rapid growth. For the steep slope–type propagation, there was an early stage incubation period, then a peak period of posting from the 24th hour to the 48th hour, followed by gradually decreasing growth. After the 114th hour,

the growth rate was almost 0. For the abrupt type, the growth rate was slow in the first 30 hours or so, and then, it increased slightly for about 2 hours. Before the 54th hour, the growth rate was almost flat, then it peaked between 54 and 72 hours and after that it gradually decreased. For the wave type, the number of tweets increased throughout the first 42 hours, then the growth rate showed a wave-shaped oscillation within 43 to 149 hours. After the 150th hour, the growth rate was almost 0. For the steep slope type, the number of tweets increased sharply over the first 6 hours or so, followed by a pause, and then by a gradual and steady increase from the 84th hour to 101st hour. After the 102nd hour, the growth rate was almost 0. For the long slope type, the growth rate of the tweets was stable and long lasting, persisting for 120 hours, after which it grew gradually and slowly.

What are the laws of propagation for the different types of violent medical incidents? We attempted to investigate this question with regard to the severity of consequences, the inducements for the incidents, and the popularity of the topic over the course of a week (in terms of the number of tweets). Our study found that the more serious the consequences of a violent incident, such as a serious injury or death, the more attention it drew on the microblog, the faster was its propagation speed, and the longer was its duration. In these cases, the propagation types were mostly of the steep slope, long slope, and zigzag types. In addition, the more serious the consequences of a violent incident, the greater popularity it exhibited on the microblog. Acts resulting from patients' dissatisfaction with treatment effects led to significantly greater popularity within a week than incidents caused by nontherapeutic effects.

Managerial Insights

Our research has important practical implications for both hospitals and governments. Workplace violence toward medical staff happens more often in China than in other countries, which provides multiple opportunities for research on the classification and prediction of incidents and their discussion on microblogs. The findings of this paper provide a new method for hospitals and health management departments in China and other countries to accurately predict communication types and rules of propagation in the discussion of violent incidents and can help relevant departments effectively manage public opinion concerning doctors and patients, ease workplace violence in the health sector, and improve doctor-patient relationships. Specifically, hospitals and health management departments should establish a public opinion monitoring system for doctor-patient relationships. When a violent incident occurs, relevant departments should pay close attention to the relevant tweets and comments and scientifically predict the communication types and rules of propagation based on the causes and severity of the consequences, so as to take effective countermeasures. In the case of incidents that lead to serious consequences, as these incidents would be quickly taken up in public discussion, hospitals and health management departments should promptly investigate the causes and announce the progress and results of the incidents. Some scholars emphasize the necessity to use both traditional media and social media in crisis response for mutual communication [10,32]. In the case of incidents that are caused by patients' dissatisfaction with treatment effects, which the public often attributes to the low technical ability of medical staff and which easily generate heated discussions and lead to negative impacts on hospitals and staff, hospitals and health management departments should investigate the causes; take responsibility; and, through official channels and in a timely manner, clarify the situation, answer questions, and assuage doubts raised by the public.

Limitations and Future Research

Our work has certain limitations that provide directions for future research. In this study, we explored the types and laws of the communication of events on the Sina microblog and analyzed the relationship among the types of communication, the laws of propagation, the causes of events, and the severity of the results. However, this study did not consider factors—such as the hospital type, the response strategies of health management departments and hospitals, and settlement results from judicial authorities—that could have an impact on the propagation of discussions of violent incidents on the microblog. We hope that future scholarly work, while attempting to confirm or refute our key findings, could also complement and expand the abovementioned issues. Future research is thus needed to select more incidents for the prediction analysis and improve the accuracy of the predictions.

Acknowledgments

This study was supported by National Social Science Foundation of China (15CGL070), Sichuan Department of Science and Technology (2019JDR0102), Sichuan Hospital Management and Development Research Center (SCYG2017-15), the Mega-projects of Sichuan Social Science Research for the 13th 5-Year Plan (SC18A008), and Xinglin Scholars project of Chengdu University of Traditional Chinese Medicine (QNXZ2018048).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Samples of violence against medical staff.
[DOCX File , 20 KB - jmir_v22i5e13294_app1.docx]

Multimedia Appendix 2
Multiple comparisons.
[DOCX File , 18 KB - jmir_v22i5e13294_app2.docx]

References

1. Zhao Y, Wangfeng M, Zhou CY, Xu W. Investigation and analysis of present situation of workplace violence at tertiary hospitals and county hospitals in Heilongjiang Province. Med and Soc 2016;29(9):35-37. [doi: 10.13723/j.yxysh.2016.09.011]
2. Driscoll RJ, Worthington KA, Hurrell JJ. Workplace assault: an emerging job stressor. Consult Psychol J Prac Res 1995;47(4):205-212. [doi: 10.1037/1061-4087.47.4.205]
3. Fang H, Zhao X, Yang H, Sun P, Li Y, Jiang K, et al. Depressive symptoms and workplace-violence-related risk factors among otorhinolaryngology nurses and physicians in Northern China: a cross-sectional study. BMJ Open 2018 Jan 27;8(1):e019514 [FREE Full text] [doi: 10.1136/bmjopen-2017-019514] [Medline: 29374676]
4. Feng C, Ji WW, Li HY. Research on the social psychological incentives and intervention strategies for the medical violent injuries. Chinese Hosp Manag 2016;36(12):66-67 [FREE Full text]
5. Hu YH, Zhang ZX, Wang GP. Analysis of inducements and countermeasures of doctor-patient conflicts based on patient experience. Chinese Hosp 2016;20(2):74-75. [doi: 10.3969/j.issn.1671-0592.2016.02.030]
6. Li LJ. Crisis management study on trinity of doctor-patient conflicts control system. Med Philos 2013;34(15):70-73.
7. Farrell GA, Bobrowski C, Bobrowski P. Scoping workplace aggression in nursing: findings from an Australian study. J Adv Nurs 2006 Sep;55(6):778-787. [doi: 10.1111/j.1365-2648.2006.03956.x] [Medline: 16925626]
8. Merecz D, Rymaszewska J, Mościcka A, Kiejna A, Jarosz-Nowak J. Violence at the workplace--a questionnaire survey of nurses. Eur Psychiatry 2006 Oct;21(7):442-450. [doi: 10.1016/j.eurpsy.2006.01.001] [Medline: 16530389]
9. Shi L, Zhang D, Zhou C, Yang L, Sun T, Hao T, et al. A cross-sectional study on the prevalence and associated risk factors for workplace violence against Chinese nurses. BMJ Open 2017 Jun 24;7(6):e013105 [FREE Full text] [doi: 10.1136/bmjopen-2016-013105] [Medline: 28647719]
10. Peterson KJ. Sociology of medicine: (2nd edn), by Rodney M. Coe. McGraw-Hill. New York, 1978. 437 pp. $16.50. Soc Sci Med B 1980;14(3):188. [doi: 10.1016/0160-7987(80)90013-7]
11. Ghaffarifar S, Ghofranipour F, Ahmadi F, Khoshbaten M. Barriers to effective doctor-patient relationship based on PRECEDE PROCEED Model. Glob J Health Sci 2015 Mar 25;7(6):24-32 [FREE Full text] [doi: 10.5539/gjhs.v7n6p24] [Medline: 26153160]
12. ILO/ICN/WHO/PSI Joint Programme on Workplace Violence in the Health Sector. Framework guidelines for addressing workplace violence in the health sector / Joint Programme on Workplace Violence in the Health Sector. Geneva: International Labour Organization; 2002.
13. AbuAlRub RF, Khalifa MF, Habbib MB. Workplace violence among Iraqi hospital nurses. J Nurs Scholarsh 2007;39(3):281-288. [doi: 10.1111/j.1547-5069.2007.00181.x] [Medline: 17760803]
14. LeBlanc MM, Kelloway EK. Predictors and outcomes of workplace violence and aggression. J Appl Psychol 2002 Jun;87(3):444-453. [doi: 10.1037/0021-9010.87.3.444] [Medline: 12090602]
15. Robinson G. Effective doctor patient communication: building bridges and bridging barriers. Can J Neurol Sci 2002 Jun;29(Suppl 2):S30-S32. [doi: 10.1017/s0317167100001918] [Medline: 12139084]
16. Duan G, Yu W, Li J. Research of characteristics and inducements of violent events to medical workers - based on 132 hot events in internet new media. Chongqing Med 2016;45(32):4535-4538. [doi: 10.3969/j.issn.1671-8348.2016.32.023]
17. Li M. New thinking on the public crisis management of hospitals in the case of malignant violent events to medical staff - taking 2012 years of successive events as examples. Theory Monthly 2013 Sep;8:108-112. [doi: 10.3969/j.issn.1004-0544.2013.08.024]
18. Xin WJ, Lai H. The dynamic mechanism of group polarization in Sina Weibo: based on content analysis of doctor slain in Wenling city. J Intell 2014 Jun:162-166.
19. Yang YF, Yu WP, Tian P. Classification and prediction of brand scandal spread on microblog based on the SOM. J Intel 2013(10):23-28. [doi: 10.3969/j.issn.1002-1965.2013.10.005]
20. Vesanto J, Alhoniemi E. Clustering of the self-organizing map. IEEE Trans Neural Netw 2000;11(3):586-600. [doi: 10.1109/72.846731] [Medline: 18249787]
21. Kohonen T. Self-Organizing Maps. New York: Springer, Berlin, Heidelberg; 2001.
22. Abrahart RJ, See L. Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments. Hydrol Process 2000;14(11-12):2157-2172. [doi: 10.1002/1099-1085(20000815/30)14:11/12<2157::aid-hyp57>3.0.co;2-s]
23. Furundzic D. Application example of neural networks for time series analysis: Rainfall–runoff modeling. Signal Process 1998;64(3):383-396. [doi: 10.1016/s0165-1684(97)00203-x]
24. Sid O, Menaa M, Hamdani S, Touhami O, Ibtiouen R. Self-Organizing Map Approach for Classification of Electricals Rotor Faults in Induction Motors. In: Proceedings of the 2011 2nd International Conference on Electric Power and Energy Conversion Systems. 2011 Presented at: EPECS'11; November 15-17, 2011; Sharjah, United Arab Emirates. [doi: 10.1109/epecs.2011.6126845]

25. Tran LT, Knight CG, O'Neill RV, Smith ER, O'Connell M. Self-organizing maps for integrated environmental assessment of the Mid-Atlantic region. Environ Manage 2003 Jun;31(6):822-835. [doi: 10.1007/s00267-003-2917-6] [Medline: 14565701]
26. Reker D, Rodrigues T, Schneider P, Schneider G. Identifying the macromolecular targets of de novo-designed chemical entities through self-organizing map consensus. Proc Natl Acad Sci USA 2014 Mar 18;111(11):4067-4072 [FREE Full text] [doi: 10.1073/pnas.1320001111] [Medline: 24591595]
27. Huysmans J, Baesens B, Vanthienen J, van Gestel T. Failure prediction with self organizing maps. Expert Syst Appl 2006;30(3):479-487. [doi: 10.1016/j.eswa.2005.10.005]
28. Ren JH, Chen JC, Wang N. Visual Analysis of SOM Network in Fault Diagnosis. Physics Procedia 2011 Dec;22:333-338 [FREE Full text] [doi: 10.1016/j.phpro.2011.11.052]
29. Mangiameli P, Chen S, West D. A comparison of SOM neural network and hierarchical clustering methods. Eur J Operational Res 1996 Sep 6;93(2):402-417 [FREE Full text] [doi: 10.1016/0377-2217(96)00038-0]
30. Costa IG, de Carvalho FD, de Souto MC. Comparative analysis of clustering methods for gene expression time course data. Genet Mol Biol 2004;27(4):623-631 [FREE Full text] [doi: 10.1590/s1415-47572004000400025]
31. Afolabi MO, Olude O. Predicting Stock Prices Using a Hybrid Kohonen Self Organizing Map (SOM). In: Proceedings of the 2007 40th Annual Hawaii International Conference on System Sciences. 2007 Presented at: HICSS'07; January 3-6, 2007; Waikoloa, HI, USA. [doi: 10.1109/hicss.2007.441]
32. Austin L, Fisher LB, Jin Y. How audiences seek out crisis information: exploring the social-mediated crisis communication model. J Appl Commun Res 2012;40(2):188-207. [doi: 10.1080/00909882.2012.654498]

Abbreviations

ANOVA: analysis of variance
RMSE: root-mean-square error
SOM: self-organizing map
SSE: sum of squares due to error

Edited by G Eysenbach; submitted 04.01.19; peer-reviewed by L Akers, Z Hong, Y Yang; comments to author 27.04.19; revised version received 30.12.19; accepted 07.02.20; published 26.05.20.

Please cite as:
Duan G, Liao X, Yu W, Li G
Classification and Prediction of Violence Against Chinese Medical Staff on the Sina Microblog Based on a Self-Organizing Map: Quantitative Study
J Med Internet Res 2020;22(5):e13294
URL: http://www.jmir.org/2020/5/e13294/
doi:10.2196/13294
PMID:32348253

Original Paper

How Geographical Isolation and Aging in Place Can Be Accommodated Through Connected Health Stakeholder Management: Qualitative Study With Focus Groups

Sonia Chien-I Chen[1*], BA, MSc, PhD; Chenglian Liu[2*], BSc, MSc, PhD; Zhenyuan Wang[3,4*], PhD; Rodney McAdam[5*], PhD; Michael Brennan[5*], PhD; Shirley Davey[5*], BSc, MSc, PhD; Teng Yuan Cheng[6*], PhD

[1]Institute of Quantitative Economics, Huaqiao University, Xiamen, China

[2]School of Computing, Neusoft Institute of Guangdong, Foshan, China

[3]Faculty of Economics and Management, East China Normal University, Shanghai, China

[4]Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai, China

[5]Ulster Business School, Ulster University, Newtownabbey, United Kingdom

[6]Nanjing Audit University, Naijing, China

[*]all authors contributed equally

Corresponding Author:
Zhenyuan Wang, PhD
Faculty of Economics and Management
East China Normal University
Room 307, S Fashang Building
No 500, Dongchuan Rd, Minhang District
Shanghai, 201108
China
Phone: 86 15800705298
Email: wangzhenyuan@dbm.ecnu.edu.cn

Abstract

Background: In remote areas, connected health (CH) is needed, but as local resources are often scarce and the purchasing power of residents is usually poor, it is a challenge to apply CH in these settings. In this study, CH is defended as a technological solution for reshaping the direction of health care to be more proactive, preventive, and precisely targeted—and thus, more effective.

Objective: The objective of this study was to explore the identity of CH stakeholders in remote areas of Taiwan and their interests and power in order to determine ideal strategies for applying CH. We aimed to explore the respective unknowns and discover insights for those facing similar issues.

Methods: Qualitative research was conducted to investigate and interpret the phenomena of the aging population in a remote setting. An exploratory approach was employed involving semistructured interviews with 22 participants from 8 remote allied case studies. The interviews explored perspectives on stakeholder arrangements, including the power and interests of stakeholders and the needs of all the parties in the ecosystem.

Results: Results were obtained from in-depth interviews and focus groups that included identifying the stakeholders of remote health and determining how they influence its practice, as well as how associated agreements bring competitive advantages. Stakeholders included people in government sectors, industrial players, academic researchers, end users, and their associates who described their perspectives on their power and interests in remote health service delivery. Specific facilitators of and barriers to effective delivery were identified. A number of themes, such as government interests and power of decision making, were corroborated across rural and remote services. These themes were broadly grouped into the disclosure of conflicts of interest, asymmetry in decision making, and data development for risk assessment.

Conclusions: This study contributes to current knowledge by exploring the features of CH in remote areas and investigating its implementation from the perspectives of stakeholder management. It offers insights into managing remote health through a CH platform, which can be used for preliminary quantitative research. Consequently, these findings could help to more effectively facilitate diverse stakeholder engagement for health information sharing and social interaction.

(J Med Internet Res 2020;22(5):e15976) doi:10.2196/15976

KEYWORDS

connected health care; remote areas; business strategy; Taiwan; population aging; knowledge sharing

Introduction

Background

The global phenomenon of an aging populations has made health care a universal issue, especially in remote areas. Connected health (CH) has been proposed as a promising solution to manage challenges arising in these remote areas. It aims to reduce isolation, enhance safety, increase efficiency, and diminish costs for remote residents [1]. It acts as an alternative solution to make "aging in place" feasible by connecting infrastructure, devices, and health care stakeholders [2]. In remote areas, CH is needed, but as local resources are often scarce and the purchasing power of residents is usually poor, it is a challenge to apply CH in this setting. In this study, remote areas were defined in 3 categories: mountain areas, isolated islands, and remote townships. In these areas, medical and care services are legally allowed to be practiced remotely according to the law and regulation of the state. There are 48 remote townships that account for 44% of the area in Taiwan, while the 0.36 million residents only account for 1.6% of the total population.

There is evidence that CH can increase access to services across a range of medical specialties without detrimental effects and improve opportunities for professional development [3]. However, most of these measures are temporary responses to government policy, and the sustainability of such services is an issue that must be addressed. Consequently, although technological interventions have improved accessibility, efficiency, and cost-effectiveness, the health status of rural residents remains a matter of concern [4]. Thus, the parameters of effective CH implementation in rural and remote contexts need to be explored to suggest strategies to manage issues appropriately.

Research suggests that the maturity of information communication technology, advances in health care, and the integration of health and social care may offer the fundamental infrastructure to boost the CH ecosystem [5]. Taiwan was selected as an ideal research area as it fulfills these requirements. Therefore, the results of this investigation may inspire those who are attempting to manage the issues of aging populations in a remote setting through a knowledge-sharing perspective.

To understand the obstacles faced in the implementation of CH in remote areas, this study explored the developmental barriers of CH in remote areas based on four representative cases in Taiwan. Geographically, CH in Taiwan has been divided into 4 parts: North, middle, South, and East. Each part has a designated hospital to support other hospitals and health centers in their areas. This study aimed to determine ideal strategies and methods by exploring the unknowns to discover insights for those facing similar issues in applying CH.

In this study, CH is defended as a technological solution that reshapes the direction of health care to be more proactive, preventive, and precisely targeted, and thus, more effective. CH provides great value in managing and preventing chronic diseases that result in tremendous burden on health care and social services. Stakeholder analysis was employed to facilitate strategy formation, as it can generate knowledge about how the characteristics of stakeholders influence decision-making processes as well as the relevant actors' behavior, intentions, and interrelations.

Theoretical Basis

The challenges of accessing appropriate health services and of recruiting and retaining staff constrain the quality of health care in remote settings [6]. With respect to a proactive approach to current problems and insights into remote health, gaps remain [7]. Possible strategies for these issues include overcoming geographic isolation and facilitating rural and local health responses. This study explored how stakeholder characteristics influence the decision-making process. The literature suggests that the need to balance conflicts of interest and the influence of organizational factors and professional support may impact the quality of health in remote areas [8,9]. Therefore, this study explored perspectives on stakeholder arrangements, including the power and interests of stakeholders and their respective needs in the ecosystem, as shown in Figure 1.

Figure 1. A simple stakeholder management model (adapted from Freeman and Reed [10]).

Stakeholder analysis is often employed to facilitate the formation of strategies, as it can generate knowledge about how the characteristics of stakeholders influence decision-making processes as well as the relevant actors' behaviors, intentions, and interrelations [8]. Courage, creativity, and a capacity to recognize opportunities for change will be key for public health advocates to create political incentives and manage political risks for leaders. Globalization may be helpful for leaders to gain awareness of the necessity of change and innovative strategies in health care [11]. Better insight into how health systems are structured and how they react over time to better adjust to health programs will also affect stakeholder management [12]. Despite considerable research on risk analysis in health politics, only little insight has been gained.

The value of stakeholder analysis in the health promotion domain is its ability to increase our understanding and capacity to participate in and contribute to health policy development by mapping the relevant actors. Once insights into their position and connection in networks are discovered, strategic planning can then be proposed [13].

General agreement among stakeholders may help the health information system advance in meeting the objectives of improving the governance, efficiency, user and provider satisfaction, and long-term fiscal sustainability of the health care system [14,15]. A stakeholder analysis framework is often used to highlight how key relationships with stakeholders may change with the perceived credibility of the organizational leaders and the legitimacy of their actions, as CH is a disruptive innovation in health care services. It is fascinating to explore how small rural health centers and a well-established, federally qualified community health center contribute to managing the risks of health care [16]. The concept of stakeholder management can help us to understand the tasks and structural changes via the legislation of decision makers [10]. Considerable systematic planning for health promotion that addresses people's lives, work, and leisure can optimize health interventions for specific contextual contingencies and target crucial factors in the organizational context that influence behavior [17].

The rise of the patient-centered concept has increased the complexity of the health care system, which motivates partnerships between the public sector and private actors in stakeholder management [18]. The literature indicates that a situation with a multitude of actors with diverse interests suggests a loosely connected network, for those who attempt to influence policy will need to work with international-level, federal-level, and regional-level actors because they play an important role in bridging and connecting the decentralized regional-level and local-level actors as well as in initiating policy engagement and change [19,20]. Multicriteria decision analysis can add value to the strategic decision-making process in health technology assessment through systematic reviews, as it is more focused on how to engage stakeholders than to explain how to develop the algorithms and methodologies [21]. Analyzing routinely captured health information and giving feedback to clinical staff have been proposed to deliver better outcomes for patients and communities in the CH program [22].

The attempt to identify factors affecting the availability, accessibility, and coordination of services serves to develop and implement culturally sensitive service delivery in remote health care settings. Findings could inform recommendations for the provision of health services to contribute to the broader knowledge of rural and remote health service provision [23]. Access to allied health services is always an issue for people with disabilities living in rural and remote areas. Evidence indicates that CH is a valid option to provide those with disabilities alternative options to receive health care services [24]. Investigation of frail older adults and their stakeholders suggests that an integrated system with a care coordinator to improve connections between services and patients is urgently needed. It is a must to reduce bureaucracy and increase the timeliness of treatment and care. Measures to improve access to health and social care systems for pre-frail and frail patients, as well as their caregivers, must be considered [25]. Also, CH is beneficial for patients with chronic conditions as well as for the frail, by improving access to health care services and allowing them to be monitored at home [26].

Strengthening the rural deployment of stakeholders may attract and motivate graduates to work in rural areas, although the factors influencing recruitment of health professionals are varied [27]. If individual knowledge of successful aging and the associated economic outcomes, such as financial planning for retirement wellbeing, are to be harnessed, stakeholder management can help put pressure on the state to improve the current health insurance system, making it possible to offer a universal social pension that prioritizes people deprived of

income due to a disability, severely debilitating disease, or lost work opportunity during their younger years [28].

Research Questions

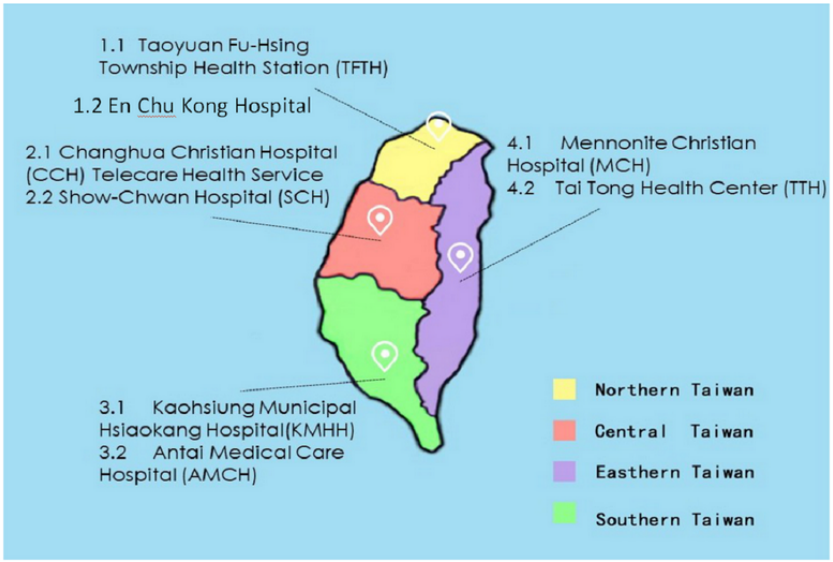
This study aimed to determine how geographical isolation and aging in place can be accommodated through CH stakeholder management based on the knowledge of the stakeholders' identities and their interests and power in remote areas of Taiwan.

Methods

An exploratory approach was conducted involving semistructured interviews with 22 participants from 8 remote allied case studies [29]. The interviews explored perspectives on stakeholder arrangements, including the power and interests of stakeholders and the needs of all the parties in the ecosystem [30,31]. To manage health challenges for the aging population, the Taiwanese government divided Taiwan into 4 parts to conduct a pilot study of CH in 1996 (Figure 2). This provided the foundation for the current infrastructure of CH and smart health in Taiwan. Therefore, this study organized research according to this foundation. An exploratory approach was conducted involving semistructured, in-depth interviews with 22 participants in four remote allied case studies: Northern Taiwan, Central Taiwan, Southern Taiwan, and Eastern Taiwan (Figure 2). Northern Taiwan is represented by 1.1 Taoyuan Fu Hsing Township Health Station and 1.2 En Chu Kong Hospital; Central Taiwan by 2.1 Changhua Christian Hospital Telecare Health Service and 2.2 Show-Chwan Hospital; Southern Taiwan by 3.1 Kaohsiung Municipal Hsiaokang Hospital and 3.2 Antai Medical Care Hospital; and Eastern Taiwan by 4.1 Mennonite Christian Hospital and 4.2 Tai Tong Health Center.

This study investigated how geographical isolation and aging in place can be accommodated through CH stakeholder management with reference to, but not limited by, the interview questions outlined in Textbox 1, based on a qualitative method with focus groups.

Figure 2. Remote health settings in Northern, Central, Southern, and Eastern Taiwan.



Textbox 1. Interview questions used with the focus groups.

1. Who are your stakeholders?
2. How do they influence your business?
3. Do you have cooperative agreements with other organizations?
4. Are these agreements offer you competitive advantages?

Ethics

Ethical approval was obtained from the Ethics Committees of Ulster University (reference no RG3 RMcAdam2). Participants gave verbal consent according to the ethical guidelines.

Participant Selection

The authors selected 8 allied case studies from 4 geographical areas covering all remote areas in Taiwan based on a literature review and the foundation of the governmental pilot schemes referred to in Figure 2. Participants and institutions were chosen accordingly. The participants included those who were responsible for CH projects; patients were excluded due to ethical concerns. Residence in the four administrative areas based on geographic and rural location was the main criterion for selecting participants, and experience with CH was the second selection criterion. These sampling criteria were chosen in order to include the inputs of major and significant participants with various levels of rural experience and management. Consequently, 8 health institutions with 22 participants that represented all remote areas of Taiwan were selected, as shown in Figure 2 and Table 1.

Table 1. Allied health professionals participating in the study.

Geographic area and health care institutions	Role and number of participants	Age (years), range
1. North		
1.1 Taoyuan Fu Hsing Township Health Station	1 GP[a], 1 nurse, 1 civil servant	40-50 (GP and civil servant), 30-40 (nurse)
1.2 En Chu Kong Hospital	1 nurse	30-40
2. Central		
2.1 Changhua Christian Hospital (CCH) Telecare Health Service	4 nurses	30-40
2.2 Show-Chwan Hospital	4 nurses	30-40
3. South		
3.1 Kaohsiung Municipal Hsiaokang Hospital	1 GP, 1 pharmacist	30-40
3.2 Antai Medical Care Hospital	1 nurse, 2 GPs	30-40 (nurse), 40-50 (GPs)
4. East		
4.1 Mennonite Christian Hospital	1 nurse, 1 IT[b] director	40-50
4.2 Tai-Tong Health Centre	1 GP, 2 social workers	25-45

[a]GP: general physician.

[b]IT: information technology.

Data Collection

Following informed consent, researchers conducted face-to-face interviews with 22 participants lasting from 30 minutes to 2 hours. To ensure the significance, validity, and rigor of this qualitative research, several research strategies were conducted regarding the quality of data collection. These protocols include the use of the agile method to allow interviewees' opinions to be validated and exchanged through a semistructured interview manner. Any differences between the interviews were discussed until consensus was reached. Interviews were audiotaped and transcribed verbatim by an independent typist with subsequent validation by the researcher. Meaningful quotations were adduced to represent important themes. To ensure participant confidentiality, data were de-identified before processing.

Data Analysis

In this study, data were categorized using stakeholder and thematic analyses. Stakeholder analysis is conducted according to the degree of the stakeholders' ability to affect health care providers' objectives and health care providers' ability to affect stakeholders' objectives. It follows 4 steps to investigate how health institutions influence stakeholders for their interests: (1) identify stakeholders, (2) assess their interests and influence, (3) develop a communication management plan, and (4) engage and influence stakeholders. In the thematic analysis, codes were derived from the data through several steps including data cleaning, data summarizing, data analysis, and data mining. The categorization of data was continually revisited and reviewed until the themes and categories used to summarize and describe the findings were verified and accurately reflected the data. A qualitative data management system, NVivo 12 (QSR International, Melbourne, Australia), was employed to manage the data throughout the process. First, data were cleaned through an integration process to merge different terms with the same meaning. Thus, CH might be called "remote health," "telehealth," or "telecare" in the interviews, and these terms were merged according to the actual meaning of the interviewees. Second, the data were summarized, clustered, and categorized based on the interviewees' meaning. Third, the data were analyzed and extracted according to the interviewees' meaning at the stage of data mining. For example, some issues were raised by interviewees, but the meanings and root causes of the issues needed to be analyzed.

Results

Participants

A total of 22 participants representing 8 CH groups that covered all the remote areas in Taiwan were involved in this study. The majority of the participants were health professionals such as nurses (12/22, 54%), pharmacists (1/22, 5%), and general physicians (5/22, 22%), accounting for more than three-quarters of the sample. In addition, social workers accounted for approximately 10% (2/22, 9%) of the sample. Information technology professionals and administrators each accounted for 5% of the participants (1/22, 5% each), as shown in Multimedia Appendix 1.

Stakeholder Analysis

Stakeholder analysis was conducted according to the degree of the health care providers' ability to affect their stakeholders' objectives and interests. Four steps of stakeholder management were used: (1) identify stakeholders, (2) assess their interests and influence, (3) develop a communication management plan, and (4) engage and influence stakeholders. Due to the complexity of the data, these steps are described in Multimedia Appendices 1-4. The results revealed the identity of the stakeholders of CH in remote areas, how they influence practices, and how associated agreements bring competitive advantages.

The stakeholders for 8 participating health facilitators were identified (Multimedia Appendix 1). These stakeholders included industrial players (software developers, hardware manufacturers, total solution providers, network providers), users and their associates (end users), government sectors, and academic researchers, which collectively constitute an ecosystem of CH. The industrial players included both software developers and total solution providers (eg, Fora Care Inc, Asus Cloud, Huede Technology, and Far EasTone Telecommunications), while some telecommunications companies fell in both categories of network providers and software developers (eg, Far EasTone Telecommunications and Chunghwa Telecom). This fact may suggest that the CH industry is a multidisciplinary domain. What is noticeable is that not all the health facilitators had a telehealth or telecare center; therefore, some had to rely on industrial players to analyze and manage collected data. Compared with public health facilitators, private health facilitators have different attitudes toward government sectors. Although all the end users were from remote areas, they could originate from mountain areas, remote townships, or isolated islands, which differentiates the nature of their health facilitators.

In Multimedia Appendix 2, the assessment of the interests of each CH stakeholder group and the influence of health facilitators is described. Government sectors are concerned with how to identify significant outcomes to increase and promote political publicity. For academic researchers, innovative topics for research and how to explore unknown and novel concepts to contribute to the body of knowledge seem to be of the utmost importance. End users are looking for user-friendly, efficient, and cost-effective solutions to manage their health care according to health facilitators. However, they may care more about their financial health than their physical health. Although most industrial players are interested in increasing sales through developing and perfecting their products and services, the concepts of software developers and hardware manufacturers will vary due to the nature of their businesses. What is noticeable is that the influence on stakeholders determines the power and strategies of health facilitators.

According to the assessment in Multimedia Appendix 2, potential strategies for health facilitators to develop a communication management plan for stakeholders are shown in Multimedia Appendix 3. First, the communication constraints of health facilitators should be determined, followed by identifying the information to be communicated. Finally, the methodology for communications is determined, to facilitate the communications. The challenges for health facilitators include finding the right people to deliver the communication. The participants indicated that information technology professionals usually do not comprehend their requests and even if they do eventually understand them, they may not be able to make relevant decisions. Moreover, it is difficult to include interdisciplinary people who can accommodate supply and demand from various areas in the CH ecosystem. These considerations constrain communication between health facilitators and their stakeholders. Although the intent to fulfill the stakeholders' interests by exchanging resources is helpful, the degree of the bargaining power of stakeholders differs to that of health facilitators. Regarding the information to be communicated, many health facilitators admit that they still struggle. It should be stressed that strategies for increasing the visibility of institutions can facilitate communication with stakeholders.

From the data presented in Multimedia Appendix 3, the benefits of increasing the profile of health facilitators are clear. Regarding the next phase, Multimedia Appendix 4 provides information about how health facilitators can engage and influence their stakeholders according to the developed communication management plan. Then, the actual plans and strategies put in use are included. Health facilitators might influence stakeholders according to their interests. The participants indicated that industrial players offer free CH samples to receive test feedback from remote areas, especially mountain areas, because the features of geographic isolation and the local authority of general physicians (GPs) could significantly contribute to the development of a considerable number of products that could be brought to the market. After health facilitators achieve a solid reputation in the CH ecosystem, many opportunities for cooperation will spontaneously appear. Although the participants suggested that end users are usually concerned about the affordability of CH products and services, GPs may play an important role in educating patients about the value of CH. Other health facilitators, such as nurses, usually offer alternative solutions for patients to consider. Overall, those who have a typical environment or samples for CH to implement have stronger bargaining power when dealing with industrial players.

As mentioned, stakeholders representing government sectors, industrial players, academic researchers, and end users and their associates. The stakeholder analysis is mapped in Figure 3.

Figure 3. Stakeholder analysis of remote health care (source: adapted from Freeman and Reed [10]).

Initially, industrial players have relatively high importance, as they possess both technologies and economic benefits. Health providers in remote areas should, therefore, manage these closely. Next, government sectors are essential as they are key for funding the remote services as well as implementing rules and regulations. Remote users and their associates are less powerful in the remote health ecosystem than the parties in the first two categories, as they are usually voluntary and have limited economic strength. They are to be kept informed of any updates. Finally, academic researchers are less influential in affecting health care providers' objectives. Thus, they play the role of serving as a reference.

> *Governmental funding is important as we have limited capability to commercialize remote health care in the initial stage…*
>
> *The main purpose of offering remote health care is to address the social equality of health care rather than earning money.*
>
> *Although remote areas lack resources, the geographic features can be ideal locations for industrial players to develop and test their products and services.*
>
> *It is essential to keep connected to the latest health information and progress as well as global trends through academic interactions.*

Full Disclosure of Conflicting Interests

Several themes, such as government interests and power of decision making were corroborated across remote services. These have been broadly grouped into the disclosure of conflicts of interest, asymmetry in decision making, and data development for risk assessment. Health care stakeholders are criticized for finding it hard to reach common ground, as they all have different interests. Industrial players tend to emphasize economic benefits rather than ethics and user privacy, while health care providers value security, ethics, and patient rights above all. Government sectors are concerned with political forces, and academic researchers care most about the social impact.

> *Government is propaganda…they only care about their political achievements.*
>
> *We care about remote residents' lives, feelings, and rights to receive care.*
>
> *The population of remote areas is low; therefore, many businesses tend to ignore residents' interests as their scale is too small to yield economic benefits.*
>
> *A successful experience building up remote health models in Taiwan and expanding them worldwide may contribute significant impacts globally.*

Although stakeholders in the remote health care ecosystem have various interests, they need to rely on each other to survive. Health providers need government's support in establishing infrastructure and initial funding; industry's technologies and services are needed in operating remote offerings alongside academia's publicity and management's inputs to increase their reputation and profile-raising. However, health providers in remote areas usually have little negotiating power in pricing and budget. Therefore, the skill of translating risks into opportunities is desirable to obtain benefits.

> *Big companies offer quality products to our telehealth center for the sake of our hospital; otherwise, they don't care much about small businesses.*
>
> *In remote areas, the infrastructure is poor, as the main telecom companies cannot gain profit from building it. It creates a challenge to practicing remote health even though it is essential for residents.*
>
> *Most of our infrastructure relies on government funding.*
>
> *Government will be happy to support us as long as the practice can show benefits to their political achievements.*
>
> *Many main industrial players have difficulty meeting customers' needs and designing user-friendly products; therefore, they come to us with free products to obtain testing samples in remote settings.*

Asymmetry (Irregularity) in Decision Making

Although remote health care focuses on serving rural residents, their capability of decision making mainly depends on health care professionals, industrial players, and government.

> *Physicians have authority over patients, so their suggestions can strongly influence patients' decisions.*
>
> *If no remote infrastructure and services from industrial players, no connected health can be set up,*

> *but unfortunately, remote residents have no word in decision making.*

As rural residents often are older, of low socioeconomic status, and poor health status and have limited purchasing power, their interests are safeguarded by government and health professionals. Fortunately, as they live in a democratic society, they can wield influence on the government through voting, which provides incentive for the government to invest in remote infrastructure – to earn political capital.

Although remote residents have little power and influence on the CH setting, their voice is often heard through GPs and other health facilitators: "Remote residents pay the same health insurance fees to the government; however, they hardly enjoy any health benefits due to poor infrastructure. It is unfair to them."

It takes effort for GPs in remote areas to fight for residents' rights. In the beginning, outcomes are limited, but little by little, their efforts have produced some fruit.

> *We have tried to contact the main telecom company; however, they show no interest in building infrastructure here as there are no economic benefits.*
>
> *Luckily, we obtained support from the Minister, so that we can have some progress in remote services.*
>
> *Thanks to the attention of the Minister, we got attention through publicity, and industrial people contacted us to discuss cooperation opportunities by offering free products for testing.*

Data Development for Risk Management

To advance the remote health decision-making process, there is a need to think strategically about how data can inform risk, as the challenges of delivering CH in remote areas can travel in two directions. The upside of this challenge may generate better outcomes, but conversely, the downsides of the outcomes can be worse than expected. Geographic isolation and poor health resources may be downsides for CH, yet strategic planning about stakeholder management may maximize opportunities for gain, which may provide entrepreneurial opportunities as companies make money by converting challenges to opportunities. What distinguishes challenges and opportunities is therefore central to business success. Some challenges are having an adverse or positive effect on an organization's profits from overseas activities. Talent that can integrate and facilitate resources to optimize the interests of CH's stakeholders is required.

> *Practicing CH in remote areas does offer many opportunities. There are many examples, including boosting the development of high technologies, offering indications for health care policy making, and shortening the gaps in health implementation.*
>
> *Our experience with CH in remote areas has not only contributed to many publications but also raised the profile of our hospital.*
>
> *The successful experiences of practicing CH may be the best advertisement and marketing strategy for our hospital.*
>
> *Taiwan's geographic features are significant in practicing CH, as it is a representative case study…if CH devices can serve in such a challenging setting as Taiwan's remote areas, there will be no issues in using it all over the world.*

To move forward, a new decision paradigm is needed with the flexibility to consider new insights and scientific information. This approach would not create an environment in which the discussion of risk based on the information is avoided. Currently, although most state and federal regulations are not designed to protect individuals, they protect the public without defining what the public is or how many individuals constitute the public. As part of a new paradigm, researchers and policymakers should carefully consider whether current federal regulations are in fact designed to adequately protect individuals, especially those in vulnerable subpopulations. Any procedural change is an opportunity to engage stakeholders on how these regulations are structured to address these populations and in what contexts. Finally, the paradigm should incorporate evaluation in the decision-making process, as assessing the impact of a decision is vital to the success of future decision making.

Discussion

Principal Findings

This study aimed to identify how geographical isolation and aging in place can be accommodated through CH stakeholder management based on identification of the stakeholders and their interests. We determined the interests of stakeholders and strategies that health facilitators can employ. Industrial players look for participants to test their products and a CH environment to comprehensively develop products for better sales. Governments seek political achievement, and academic researchers pursue interesting topics and data to have an impact society. Offering benefits that can fulfill stakeholders' interests will help to overcome the challenge of resource shortages in remote areas. However, it is not a "one-size-fits-all" solution. Health facilitators should recognize their competitive advantages to differentiate their strengths from those of others. Once their uniqueness can meet their supplier's need, they are able to exchange benefits to address their shortages.

Compared with current and past literature, our findings offer further insights into stakeholder engagement. The knowledge foundation of this study was developed from a current understanding of stakeholders' perspectives from the literature. The importance of the identification of stakeholders in electronic health has been reported [32,33]. Based on this, the concept of the co-creation and co-design of a health facility with stakeholders has been discussed [34,35]. Perspectives on a patient-centric model include an increase in technology adoption [36]. Strategies for successful implementation of technology for aging in place have been studied [37]. Person-centeredness, clinician acceptability, and informatics feasibility have been achieved and ensured through technology applications [35].

Regarding the research context, most previous literature has focused on the health care system in Western or developed countries where remote areas are numerous and internet infrastructures are better developed [4,38-41]. These countries

encounter the challenges of an aging population earlier than developing or underdeveloped countries. In contrast, stakeholder perspectives have not been extensively studied in some regions in Asia that are experiencing a similarly aging population and are interested in developing CH. Taiwan can be seen as a typical case; its rate of aging and degree of advanced technology are as high as those in Western countries [3,42-44]. Additionally, Taiwan has representative remote areas and developed health care performance [3,43]. Therefore, stakeholder engagement in CH should be of interest to improve the challenges with resource shortages in remote areas. Moreover, Taiwan has an integrated health and care system, which can facilitate health information sharing and social interaction more effectively [45].

Possible Bias in Participant Selection

Bias should be prevented to ensure the validity and value of research. However, it is difficult to avoid bias due to its complex factors. Bias can occur from the research environment, participants, and even the researchers themselves. Some researchers may intentionally influence participants to obtain the results expected. Moreover, some bias from researchers occurs unconsciously, which makes it more difficult to prevent. This phenomenon is especially present in qualitative research because qualitative research relies more on the experience and judgment of the researcher. Also, the type of data collected is subjective and unique to the person or situation. Therefore, it is much harder to avoid bias in qualitative research than in quantitative research.

Recognizing that bias exists in all research may be a good start to avoid bias. Then, the researcher should be informed about potential bias so he or she can avoid bias as much as possible. Before the research has been conducted, design bias can occur. Omission bias from selection or sampling may also occur during the research. If the study is not well planned, inclusive bias can occur when researchers look for quick solutions and convenient options.

To avoid bias, the researcher first checked the guidance for qualitative research. Second, the interview plan was discussed with the supervisor at various times to maintain the objectiveness of the interview questions. Indirect and open-ended questions were asked to allow information to flow more freely. Third, all the interviews were recorded and analyzed without personal preference. Finally, independent reviewers were invited to review the research to help maintain objectivity in the research.

Validity

In terms of the validity of the data analysis, this study followed rigorous qualitative research principles to ensure the credibility, transferability, and reliability of the data [29]. The triangulation approach was employed as a testing method to examine the validity of research [30,31]. It helps to capture different dimensions of the same phenomenon. In this study, research data triangulation and investigator triangulation were adopted to test the validity and reliability of research. In the research data triangulation, the interview questionnaires were discussed using the Delphi method with 8 focus groups and analyzed using NVivo software. In the investigator triangulation, multiple investigators, such as health professionals and experts, were invited to evaluate and review the data for selective perception and illumination blind spots in an interpretive manner to understand multiple ways of seeing the data.

Strengths

This study addressed a topic that had previously not been well studied so that it may offer insight for future projects on CH in rural areas. Different from other case studies in which the research focused mainly on developed countries, this study considered a region between developed countries and developing countries in order to discover methods for accommodating the resource shortage in remote areas. This study considered Taiwanese CH as a case study, as it meets all the CH preconditions and essential requirements. Not only are advanced technology and medicine present but the ecosystem for boosting CH is also complete and comprehensive in Taiwan. Recently, Taiwan integrated social care with its health department, to become a health and social welfare department. Chronic conditions are prevalent due to an aging demographic. Geographically, Taiwan has populations in urban areas and in many remote areas and isolated islands. Moreover, it is a mixture of public and private health care systems. These features suggest that if CH practice can be successfully applied in Taiwan, it is likely also to suit other countries around the world.

Limitations

A limitation of this study is that only one case study approach was used, making it difficult to generalize the findings due to a limited sample size. Generic principles and popular interpretations are not easily formed because the purpose of the study was to analyze problems rather than obtain summarized or statistical data. Consequently, certain inductions might be arbitrary and subjective. In addition, technical limitations and researcher bias may be an issue because a standard data analysis method was not used; therefore, the data interpretation and presentation of evidence are influenced by the researcher's choice, which may affect the results of the study. Further generalization is likely to require additional case studies for different potential contexts, which could include diverse case studies that may show cultural differences and influences.

Future Research

The outcomes of this qualitative research can serve as preliminary quantitative research on the degree of the influence of health facilitators on stakeholders' interests. Alternatively, the optimization of two or multiple objectives using a quantitative method can be applied to discover the optimal interests of all stakeholders in future work.

Conclusions

This study contributes to current knowledge by exploring the features of CH in remote areas and investigating its implementation from the perspectives of stakeholder management. Methods to accommodate geographic isolation and aging in place through CH stakeholder management are discussed based on the identification of the stakeholders and their interests and power in remote areas of Taiwan. The results offer insights for managing remote health through a CH platform, which can be used as preliminary quantitative research. Consequently, these findings could help more effectively

facilitate diverse stakeholder engagement for health information sharing and social interaction.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (NSFC, 71672060, 71372081), Ulster University, and SC Limited Company. The APC was funded by the Huaqiao University High-level Talent Research Project and The Straits Postdoctoral Fellowship Funding Program in 2018 (Fujing Financial office #2017 32/605 50X18137). We would like to thank the financial support from the Chinese National Social Science Foundation's major bidding project, "The study on theoretical Framework, International Practice and China's Paths of Finance Serving Aged" (grant #17ZDA072). All authors contributed equally.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Connected health stakeholder identification.
[DOCX File , 16 KB - jmir_v22i5e15976_app1.docx]

Multimedia Appendix 2
Connected health stakeholders' interests and influence assessment.
[DOCX File , 15 KB - jmir_v22i5e15976_app2.docx]

Multimedia Appendix 3
Connected health stakeholders' communication management development.
[DOCX File , 16 KB - jmir_v22i5e15976_app3.docx]

Multimedia Appendix 4
Stakeholders' engagement and influence.
[DOCX File , 16 KB - jmir_v22i5e15976_app4.docx]

Multimedia Appendix 5
Interview questions template.
[DOCX File , 12 KB - jmir_v22i5e15976_app5.docx]

References

1. Blount M, Batra VM, Capella AN, Ebling MR, Jerome WF, Martin SM, et al. Remote health-care monitoring using Personal Care Connect. IBM Syst. J 2007;46(1):95-113. [doi: 10.1147/sj.461.0095]
2. Bourke L, Humphreys JS, Wakerman J, Taylor J. Understanding rural and remote health: a framework for analysis in Australia. Health Place 2012 May;18(3):496-503. [doi: 10.1016/j.healthplace.2012.02.009] [Medline: 22418016]
3. Chen SC. Technological Health Intervention in Population Aging to Assist People to Work Smarter not Harder: Qualitative Study. J Med Internet Res 2018 Jan 04;20(1):e3 [FREE Full text] [doi: 10.2196/jmir.8977] [Medline: 29301736]
4. Casey M, Hayes PS, Heaney D, Dowie L, Ólaighin G, Matero M, et al. Implementing transnational telemedicine solutions: a connected health project in rural and remote areas of six Northern Periphery countries Series on European collaborative projects. Eur J Gen Pract 2013 Mar 22;19(1):52-58. [doi: 10.3109/13814788.2012.761440] [Medline: 23432039]
5. Ducat W, Martin P, Kumar S, Burge V, Abernathy L. Oceans apart, yet connected: Findings from a qualitative study on professional supervision in rural and remote allied health services. Aust J Rural Health 2016 Feb 05;24(1):29-35 [FREE Full text] [doi: 10.1111/ajr.12192] [Medline: 26052949]
6. Richmond CA, Ross NA. The determinants of First Nation and Inuit health: a critical population health approach. Health Place 2009 Jun;15(2):403-411. [doi: 10.1016/j.healthplace.2008.07.004] [Medline: 18760954]
7. Saurman E, Lyle D, Perkins D, Roberts R. Successful provision of emergency mental health care to rural and remote New South Wales: an evaluation of the Mental Health Emergency Care–Rural Access Program. Aust. Health Review 2014;38(1):58. [doi: 10.1071/ah13050]
8. Brugha R, Varvasovszky Z. Stakeholder analysis: a review. Health Policy Plan 2000 Sep;15(3):239-246 [FREE Full text] [doi: 10.1093/heapol/15.3.239] [Medline: 11012397]
9. Reed MS. Stakeholder participation for environmental management: A literature review. Biological Conservation 2008 Oct;141(10):2417-2431. [doi: 10.1016/j.biocon.2008.07.014]

10. Freeman RE, Reed DL. Stockholders and Stakeholders: A New Perspective on Corporate Governance. California Management Review 1983 Apr;25(3):88-106. [doi: 10.2307/41165018]
11. Reich MR. The Politics of Reforming Health Policies. Promotion & Education 2018 Jun 04;9(4):138-142. [doi: 10.1177/175797590200900401]
12. Blanchet K, James P. How to do (or not to do) ... a social network analysis in health systems research. Health Policy Plan 2012 Aug 12;27(5):438-446. [doi: 10.1093/heapol/czr055] [Medline: 21840934]
13. Hoeijmakers M, De LE, Kenis P, De VN. Local health policy development processes in the Netherlands: an expanded toolbox for health promotion. Health Promotion International Apr 25 2007 Jun;22(2):112-121 [FREE Full text] [doi: 10.1093/heapro/dam009]
14. Akinci F, Mollahaliloğlu S, Gürsöz H, Oğücü F. Assessment of the Turkish health care system reforms: a stakeholder analysis. Health Policy 2012 Sep;107(1):21-30. [doi: 10.1016/j.healthpol.2012.05.002] [Medline: 22652336]
15. Walt G, Shiffman J, Schneider H, Murray SF, Brugha R, Gilson L. 'Doing' health policy analysis: methodological and conceptual reflections and challenges. Health Policy Plan 2008 Sep;23(5):308-317 [FREE Full text] [doi: 10.1093/heapol/czn024] [Medline: 18701552]
16. Savage G, Dunkin J, Ford D. Responding to a crisis: a stakeholder analysis of community health organizations. J Health Hum Serv Adm 2004;26(4):383-414. [Medline: 15704640]
17. Poland B, Krupa G, McCall D. Settings for health promotion: an analytic framework to guide intervention design and implementation. Health Promot Pract 2009 Oct 06;10(4):505-516. [doi: 10.1177/1524839909341025] [Medline: 19809004]
18. Chen S, Liu C, Hu R. Fad or Trend? Rethinking the Sustainability of Connected Health. Sustainability 2020 Feb 27;12(5):1-22. [doi: 10.3390/su12051775]
19. Chen SCI, Liu C. Factors Influencing the Application of Connected Health in Remote Areas, Taiwan: A Qualitative Pilot Study. Int J Environ Res Public Health 2020 Feb 17;17(4):1282-1300 [FREE Full text] [doi: 10.3390/ijerph17041282] [Medline: 32079241]
20. Chen SCI, Hu R, McAdam R. Smart, Remote, and Targeted Health Care Facilitation Through Connected Health: Qualitative Study. J Med Internet Res 2020 Apr 28;22(4):e14201 [FREE Full text] [doi: 10.2196/14201] [Medline: 32343254]
21. Etges A, Zanotto B, da Rosa P, Polanczyk C. PP90 The Value Of Multi-Criteria Decision Analysis Use On Health Technology Decision Making Process. Int J Technol Assess Health Care 2019 Jan 03;34(S1):100-100. [doi: 10.1017/s0266462318002374]
22. Steels S, van Staa T. Evaluation protocol of the implementation of a learning healthcare system in clinical practice: the Connected Health Cities programme in the north of England. BMJ Open 2019 Jun 04;9(6):e025484 [FREE Full text] [doi: 10.1136/bmjopen-2018-025484] [Medline: 31167864]
23. Hines M, Bulkeley K, Dudley S, Cameron S, Lincoln M. Delivering Quality Allied Health Services to Children with Complex Disability via Telepractice: Lessons Learned from Four Case Studies. J Dev Phys Disabil 2019 Feb 16;31(5):593-609. [doi: 10.1007/s10882-019-09662-8]
24. Kurpas D, Gwyther H, Szwamel K, Shaw RL, D'Avanzo B, Holland CA, et al. Patient-centred access to health care: a framework analysis of the care interface for frail older adults. BMC Geriatr 2018 Nov 12;18(1):273 [FREE Full text] [doi: 10.1186/s12877-018-0960-7] [Medline: 30419817]
25. Racic M, Ivkovic N, Pavlovic J, Zuza A, Hadzivukovic N, Bozovic D, et al. Factors influencing health profession students' willingness to practice in rural regions of Bosnia and Herzegovina: a cross-sectional study. Rural Remote Health 2019 Feb 27:4717. [doi: 10.22605/rrh4717]
26. Javier JR, Conchada M, Jabar M. Anticipating Aging and Prospecting Pension for Retirement Well-being. DLSU Business & Economics Review 2019;28(2):183-198 [FREE Full text]
27. Pekmezaris R, Pascarelli B, Kozikowski A, Wolf-Klein G, Boye-Codjoe E, Jacome S, et al. Adapting Telehealth Pulmonary Rehabilitation Program to Meet the Needs of Hispanic and African-American Patients from Disparity Communities: A Community-Based Participatory Research Approach. American Thoracic Society Journal 2019 May:3737-3737 [FREE Full text]
28. Macniven R, Hunter K, Lincoln M, O'Brien C, Jeffries TL, Shein G, et al. Accessibility of Primary, Specialist, and Allied Health Services for Aboriginal People Living in Rural and Remote Communities: Protocol for a Mixed-Methods Study. JMIR Res Protoc 2019 Feb 28;8(2):e11471 [FREE Full text] [doi: 10.2196/11471] [Medline: 30816848]
29. Wilson V. Research Methods: Triangulation. EBLIP 2014 Mar 05;9(1):74. [doi: 10.18438/b8ww3x]
30. Aguinaldo J. Rethinking Validity in Qualitative Research from a Social Constructionist Perspective: From Is this Valid Research? To What Is this Research Valid for? The Qualitative Report 2004 Mar;9(1):127-136 [FREE Full text]
31. Carter N, Bryant-Lukosius D, DiCenso A, Blythe J, Neville AJ. The Use of Triangulation in Qualitative Research. Oncology Nursing Forum 2014 Aug 26;41(5):545-547. [doi: 10.1188/14.onf.545-547]
32. Petersen C, Adams SA, DeMuro PR. mHealth: Don't Forget All the Stakeholders in the Business Case. Med 2 0 2015 Dec 31;4(2):e4 [FREE Full text] [doi: 10.2196/med20.4349] [Medline: 26720310]
33. Jones R, Rogers R, Roberts J, Callaghan L, Lindsey L, Campbell J, et al. What is eHealth (5): a research agenda for eHealth through stakeholder consultation and policy context review. J Med Internet Res 2005 Nov 10;7(5):e54 [FREE Full text] [doi: 10.2196/jmir.7.5.e54] [Medline: 16403718]

34. van Woezik AFG, Braakman-Jansen LMA, Kulyk O, Siemons L, van Gemert-Pijnen JEWC. Tackling wicked problems in infection prevention and control: a guideline for co-creation with stakeholders. Antimicrob Resist Infect Control 2016 May 21;5(1):20. [doi: 10.1186/s13756-016-0119-2]
35. Kildea J, Battista J, Cabral B, Hendren L, Herrera D, Hijal T, et al. Design and Development of a Person-Centered Patient Portal Using Participatory Stakeholder Co-Design. J Med Internet Res 2019 Feb 11;21(2):e11371. [doi: 10.2196/11371]
36. Fix GM, Hogan TP, Amante DJ, McInnes DK, Nazi KM, Simon SR. Encouraging Patient Portal Use in the Patient-Centered Medical Home: Three Stakeholder Perspectives. J Med Internet Res 2016 Nov 22;18(11):e308. [doi: 10.2196/jmir.6488]
37. Peek STM, Wouters EJ, Luijkx KG, Vrijhoef HJ. What it Takes to Successfully Implement Technology for Aging in Place: Focus Groups With Stakeholders. J Med Internet Res 2016 May 03;18(5):e98 [FREE Full text] [doi: 10.2196/jmir.5253] [Medline: 27143097]
38. Mc Loughlin F, Hadgraft NT, Atkinson D, Marley JV. Aboriginal health research in the remote Kimberley: an exploration of perceptions, attitudes and concerns of stakeholders. BMC Health Serv Res 2014 Oct 26;14(1):517 [FREE Full text] [doi: 10.1186/s12913-014-0517-1] [Medline: 25343849]
39. Yazdanshenas H, Bazargan M, Jones L, Vawer M, Seto TB, Farooq S, et al. Engaging Gatekeeper-Stakeholders in Development of a Mobile Health Intervention to Improve Medication Adherence Among African American and Pacific Islander Elderly Patients With Hypertension. JMIR Mhealth Uhealth 2016 Oct 26;4(4):e116 [FREE Full text] [doi: 10.2196/mhealth.5905] [Medline: 27784651]
40. Harte R, Glynn L, Rodríguez-Molinero A, Baker PM, Scharf T, Quinlan LR, et al. A Human-Centered Design Methodology to Enhance the Usability, Human Factors, and User Experience of Connected Health Systems: A Three-Phase Methodology. JMIR Hum Factors 2017 Mar 16;4(1):e8 [FREE Full text] [doi: 10.2196/humanfactors.5443] [Medline: 28302594]
41. Hebert MA, Korabek B. Stakeholder readiness for telehomecare: implications for implementation. Telemed J E Health 2004 Mar;10(1):85-92. [doi: 10.1089/153056204773644625] [Medline: 15104920]
42. Chang H, Shaw MJ, Lai F, Ko W, Ho Y, Chen H, et al. U-Health: an example of a high-quality individualized healthcare service. Per Med 2010 Nov;7(6):677-687. [doi: 10.2217/pme.10.64] [Medline: 29788567]
43. Chen SC, Kharabsheh R. Working smart or working hard? Investigating the role of technology in healthcare innovation in Taiwan. In: Organizational Culture and Leadership's Impact on a Safety Program Change Model. Zagreb; 2017 Jul Presented at: TAKE 2017 – Theory and Applications in the Knowledge Economy; 2017/07/12-14; Zagreb p. 108-126 URL: https://tinyurl.com/y9wq4ac3
44. Liu C. Key Factors Influencing the Intention of Telecare Adoption: An Institutional Perspective. Telemedicine and e-Health 2011 May;17(4):288-293. [doi: 10.1089/tmj.2010.0184]
45. Lu Y, Wu Y, Liu J, Li J, Zhang P. Understanding Health Care Social Media Use From Different Stakeholder Perspectives: A Content Analysis of an Online Health Community. J Med Internet Res 2017 Apr 07;19(4):e109 [FREE Full text] [doi: 10.2196/jmir.7087] [Medline: 28389418]

Abbreviations

CH: connected health
GP: general physician
IT: information technology

Edited by G Eysenbach; submitted 22.08.19; peer-reviewed by J Vidal-Alaball, S Lilje, T Aslanidis, A Bashir, J Burzyńska, M Budge; comments to author 18.09.19; revised version received 27.11.19; accepted 10.04.20; published 27.05.20.

Please cite as:
Chen SCI, Liu C, Wang Z, McAdam R, Brennan M, Davey S, Cheng TY
How Geographical Isolation and Aging in Place Can Be Accommodated Through Connected Health Stakeholder Management: Qualitative Study With Focus Groups
J Med Internet Res 2020;22(5):e15976
URL: http://www.jmir.org/2020/5/e15976/
doi:10.2196/15976
PMID:32459181

Original Paper

Impact of the Price of Gifts From Patients on Physicians' Service Quality in Online Consultations: Empirical Study Based on Social Exchange Theory

Yanan Wang[1], MA; Hong Wu[1], PhD; Chenxi Xia[2], PhD; Naiji Lu[1], PhD

[1]Huazhong University of Science and Technology, Wuhan, China

[2]Jiangxi Engineering Laboratory on Radioactive Geoscience and Big Data Technology, East China University of Technology, Nanchang, Jiangxi, China

Corresponding Author:
Chenxi Xia, PhD
Jiangxi Engineering Laboratory on Radioactive Geoscience and Big Data Technology
East China University of Technology
418 Guanglan Road, Qingshanhu District
Nanchang, Jiangxi, 330013
China
Phone: 86 18963991821
Email: xcxxdy@sina.com

Abstract

Background: Gift giving from patients to physicians, which is prohibited in traditional clinical settings in China, has been found to occur in online health communities. However, there is debate on the validity of online gifts since physicians gain an economic benefit. Moreover, the potential impact of these gifts, particularly with respect to the financial value of the gift, on the online consultation service quality remains unexplored.

Objective: The aim of this study was to explore the impact of gift price on the quality of physicians' online consultation service. Insight into this impact is expected to help resolve existing debate on the appropriateness of the gift-giving practice in online consultations.

Methods: A dataset of 141 physicians and 4249 physician-patient interactions was collected from the Good Physician Online website, which is the largest online consultation platform in China. Based on social exchange theory, we investigated how gift price affects the quality of physicians' online consultation service and how this impact changes according to the physician's service price and number of all gifts received. Manual annotation was used to identify the information support paragraphs and emotional support paragraphs in the answers of physicians. The quality of the information support paragraphs, rather than the complete answer, was used to test the robustness of our model.

Results: Gift price had a positive impact on the quality of physicians' online consultation service (β=4.941, P<.01). This impact was negatively mediated by both the physician's service price (β=–9.245, P<.001) and the total number of gifts they received (β=–5.080, P<.001).

Conclusions: Gift price has a positive impact on physicians' online behavior, although the impact varies among physicians.

(J Med Internet Res 2020;22(5):e15685) doi:10.2196/15685

KEYWORDS

gift giving; gift price; service price; online consultation service quality; information support; emotional support; online health communities

Introduction

Background

Gift exchange, including giving, receiving, and reciprocating, is generally defined as the circulation of goods to promote ties and bonds between individuals [1]. Generally, gift giving has benefits for emotional expression and relationship building in interpersonal communication [2]. However, gifts exchanged between physicians and patients have been viewed as a form of illegal and unethical payment [3], which can potentially weaken the trust and deteriorate these relationships [4]. Hence, it has been forbidden for physicians to receive gifts from patients in China [4].

However, the practice of online gift giving has emerged with the development of online health communities. Online gifts were initially launched to provide patients with a channel to express gratitude to physicians, and debates about the validity of gifts in online health communities have sparked given the potential for bringing financial benefits to physicians. Some patients consider that online gift giving in online health communities represents a form of extortion with a bribe to a certain degree [5,6], whereas users of some online platforms regard it as a form of respect by acknowledging the knowledge labor of physicians [7]. In addition, physicians express different attitudes with respect to receiving online gifts. Approximately 70% of physicians stated that online gift giving is reasonable, 25% stated that it may make people uncomfortable, and 3% disapproved of this practice overall [7]. The root of these debates is the possible impact of the financial value of the gift on the quality of physicians' online consultation service, which remains unclear.

In addition to these debates, online gifts have also attracted the attention of some scholars. Zhao et al [8] studied physician-patient interactions on the online health community Good Physician Online and confirmed that online gifts from patients could improve physicians' online response rate. However, this study did not delve into the possible effects of the gift price. Gift giving, as a form of social exchange [9], follows the rules of social exchange theory (SET). Therefore, the aim of this study was to use SET to empirically explore how patients' gift-giving behaviors affect physicians' online consultation service quality from the perspective of gift price, which could help to resolve existing debates.

The deprivation-satisfaction proposition [10] in SET, which is based on the marginal diminishing effect [11] in economics, holds that the more people that receive a certain reward, the lower the perceived value of subsequent or similar rewards will be. Therefore, we considered that the total number of gifts a physician receives may have a negative moderating impact on the relationship between gift price and the physicians' online consultation service quality. Finally, comparison level theory, also included in SET, holds that individuals will consider an acceptable outcome at a comparative level according to their previous experiences, which could affect their perception and judgment of subsequent outcomes. Similarly, the physicians' service price might be considered the comparison level, which could affect their perception of the gift price. Thereby, we surmised that a physician's service price might also have a negative moderating effect on the impact of gift price on the quality of their online consultation service. More specifically, this study addressed the following three research questions: (1) how does gift price affect the quality of physicians' online consultation service? (2) how does the impact of gift price on the quality of physicians' online consultation service differ for physicians with different service prices? and (3) how does the impact of gift price on the quality of physicians' online consultation service differ for physicians receiving different numbers of online gifts?

Related Research

SET is one of the most influential conceptual paradigms for understanding workplace behavior [12]. This theory can be traced back to at least the 1920s [13,14], bridging such disciplines as anthropology [15,16], social psychology [10,17,18], and sociology [19]. Although different views of social exchange have emerged, it is widely accepted that social theory involves a series of interactions that generate obligations [20], and these interactions are usually perceived as interdependent and contingent on the actions of another person [19].

Along with rapid development of the internet, online communities have become popular platforms for carrying out routine activities of daily life. Online interaction is an important part of the online community. Such interactions represent social exchanges between participants, which occur in a network context [21]. Therefore, many human behaviors on online communities have been studied in the context of SET. For example, some scholars have focused on purchase intention [22,23], online trust [24,25], and the information contribution intention of consumers in the marketing field [26], along with aspects of self-disclosure [27], reciprocal intention of knowledge [28], and member commitment [29].

Social exchange behaviors in online health communities mainly occur between physicians and patients. Researchers have traditionally explained user behaviors in these communities based on SET, such as the knowledge sharing intention [30], influence factors of providing social support [31], and the motivation of physicians to participate in online health communities [32]. Moreover, gift giving is a form of social exchange, and some researchers have also studied the influencing factors of gift-giving frequency in social network services based on SET [33]. However, less attention has been paid to gift-giving behaviors in online health communities. Therefore, the aim of the present study was to explore gift-giving behaviors in online health communities based on SET.

Hypotheses Development and Research Model

Gift Price and Quality of Physicians' Online Consultation Service

Reciprocity, an important prerequisite for continuous exchange, refers to the fact that one has an obligation to return another's favor [17,34]. Moreover, reciprocity is not only the basis of gift exchange but is also the impetus for the flow of gifts [35]. In addition, the law of equality sets norms for social exchange according to the proportion of pay and return in the process [19]. Generally, interpersonal relationships can only be maintained based on the principles of reciprocity and equality [10].

Individuals are generally motivated to interact with others when they expect positive results [36]. Likewise, patients' gift-giving behaviors in online health communities reflect their expectations for positive results such as higher quality of consultation services and more harmonious physician-patient relationships. However, a physician's participation in an online health community is a social exchange process of professional capital for social and economic returns [32]. Therefore, a social exchange process

could form through patients giving gifts and physicians reciprocating. According to the principles of reciprocity and equality, a physician may provide an online consultation service with relatively higher quality according to the price of the gift that they received. Therefore, we proposed hypothesis 1: gift price has a positive impact on the quality of physicians' online consultation services.

Moderating Role of Gift Number

The marginal diminishing effect principle [37] posits that the value of a commodity depends on the perceived utility value by people. From this theory, the value of the marginal units of a commodity will decrease with an increase in the number of this commodity. Moreover, Homans' [10] deprivation-satisfaction proposition, according to the marginal diminishing effect, holds that the more people that receive a certain reward, the lower the perceived value of subsequent or similar rewards will be.

The reason for this marginal diminishing effect is that from the perspectives of physiology and psychology, both satisfaction and stimulation will decline with each newly added consumption unit [38]. Similarly, early gifts can bring more pleasure and stimulus to physicians. However, as the number of gifts gradually increases, physicians may regard the gift-giving behaviors of patients as a normal part of their interactions. Thus, we hypothesized that physicians' perceived value of gifts would gradually decline as the number of gifts they receive increases. Thus, hypothesis 2 is that the number of gifts that physicians receive negatively moderates the impact of gift price on the quality of physicians' online consultation service.

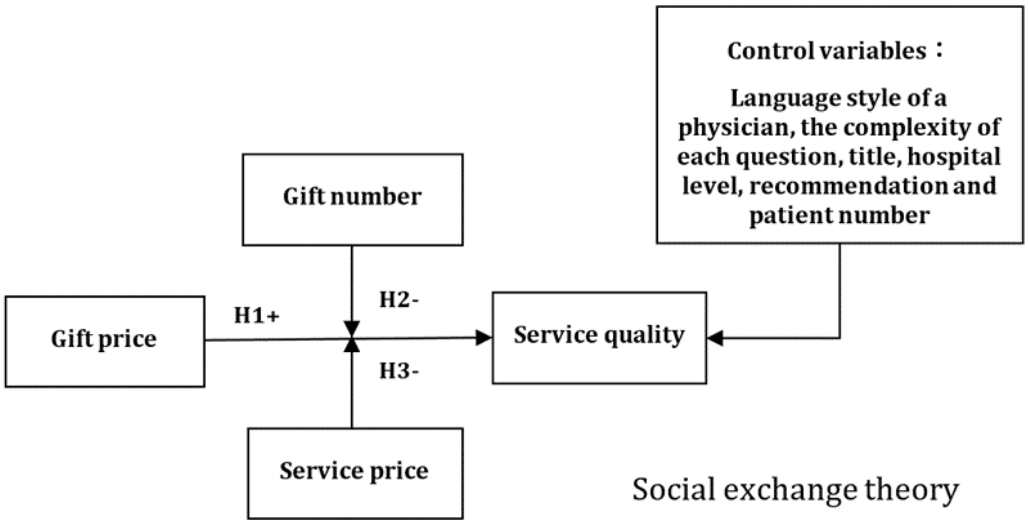
Moderating Role of Physicians' Service Price

Comparison level theory [18] was initially developed to study people's assessments of interpersonal relationships, indicating that people will assess a relationship by comparing it with a known comparison level, which could be determined from one's previous relationships or similar relationships [18]. Specifically, whether a relationship is attractive or satisfactory can be determined by comparing it with the comparison level.

In online consultations, physicians may first receive a consulting fee and thus form a comparison level. When they receive a gift during this process, they may compare the price of the gift with the comparison level. Therefore, a physician's satisfaction with the gift price would be lower when their service price is higher. Furthermore, the positive impact of gift price on the quality of the physician's online consultation service will be weaker when their service price is higher, and vice versa, leading to hypothesis 3: service price negatively moderates the relationship between gift price and the quality of physicians' online consultation services.

Based on the above hypotheses, we constructed the research model that is summarized in Figure 1.

Figure 1. Research model and hypotheses based on social exchange theory. H1: hypothesis 1, gift price has a positive impact on the quality of physicians' online consultation services; H2: hypothesis 2, the number of gifts that physicians receive negatively moderates the impact of gift price on the quality of physicians' online consultation service; H3: hypothesis 3, service price negatively moderates the relationship between gift price and the quality of physicians' online consultation services.



Methods

Research Context

Founded in 2006, Good Physician Online (haodf.com) is the largest online consultation service platform in China. By December 2018, 580,000 physicians from 9379 regular hospitals were participating on the website [39]. Good Physician Online is also the most popular platform for online gift behaviors, and the number of gifts that physicians receive can reach up to several thousand. Hence, we selected this website as the data source for our research.

The main online consultation services provided by physicians in Good Physician Online include written consultation, telephone consultation, and appointment service. In a written consultation, patients need to first pay before they are able to communicate with physicians through text or pictures if necessary. In addition, patients can also purchase online gifts for physicians on the platform to express emotions. Representative examples of the physician-patient interaction and gift-giving behavior are shown in Figure 2.

Figure 2. Examples of physician-patient interactions representing gift-giving behavior.

Data Collection

To eliminate the effects related to seeking consultations for different diseases, we collected data for physicians who only treat diabetes for the following reasons. First, diabetes is one of the most common chronic diseases. According to the 8th edition of the Global Diabetes Map released by the International Diabetes Federation, approximately 425 million adults worldwide were suffering from diabetes in 2017, including 114 million in China, ranking first in the world [40]. Second, diabetes, as a typical chronic noncommunicable disease, has the characteristics of a long onset, complicated etiology, and high difficulty to cure. Therefore, patients with diabetes are more likely to establish long-term and stable online interactions with physicians.

To empirically validate our research model, we developed a crawler targeting physicians who mainly treat patients with diabetes on the Good Physician Online website from October 12 to 13, 2018. The dataset included the personal information of the physicians and the physician-patient text dialogs during consultations. Finally, 141 physicians were included in the dataset, and the number of gifts that were exchanged in the consultations was 4249.

Variables and Empirical Model

Medical service quality has long been a primary focus of scholars in the medical field, including in an online health context. However, a standardized scientific measure of medical service quality in online contexts has not yet been developed. Many studies have focused on the assessment of online health information quality and thus proposed some mature evaluation methods and indicators [41-43]. Based on these methods, some scholars have established the evaluation criteria of online medical consultation services from the perspectives of linguistic characteristics and information quality [44]. However, such evaluation criteria are generally used in assessments of expert consultation because medical information is considered to be professional. Therefore, such evaluation criteria were not considered suitable for the large sample data used in our study.

In the field of electronic commerce, longer reviews often include more product details, which can reduce the uncertainty of product quality [45]. Likewise, a physician's answer is the main determinant of his or her service quality because online consultation services in online health communities are mainly carried out in the form of patients providing a question and physicians providing an answer. Furthermore, the average word count of a physician's answer in a dialog represents each answer's information content in the dialog. Therefore, we used the average word count to measure the quality of physicians'

online consultation services (Phy_Answer) as the dependent variable of our model.

The independent variable was the gift price (Gift_Price): Good Physician Online provides gifts with different prices, in the range of 1 to 200 yuan, for patients to choose from.

In addition, the following two moderating variables were included in the model: (1) service price (Service_Price), as patients need to pay for physicians' online consultation services, and (2) gift number (Gift_Number), which represents the total number of online gifts that physicians received.

Finally, we introduced the following control variables. The first was physicians' language styles (Phy_Style), which was based on the average word count of each physician answer to control for differences caused by physicians' different language styles. The second control variable was the complexity of patient questions (Ques_comp), since the severity of diseases could affect the complexity of questions, which was also controlled based on the average word count of each patient question in a dialog. The title (Title) was also controlled for in the model. Physicians' titles represent their professional ability and offline reputation, and there are three types physicians that participate in Good Physician Online: chief physicians, associate chief physicians, and attending physicians. Since most of the physicians using the website are chief physicians, we combined the other two categories and used one dummy variable to measure physicians' titles (see Table 1). The hospital level (Hospital_Level) was further controlled, which is evaluated by government health departments, and physicians working at hospitals of different levels have access to different medical resources. In Good Physician Online, hospital levels are classified as level A, B or C, with A being the highest quality. However, most of the hospitals that physicians on the website belong to are level A hospitals; therefore, we combined the other two categories and used hospital level as a dummy variable (Table 1). Recommendation was another control variable, which ranges along a scale of 0-5 with 5 being the best, and represents the comprehensive popularity of a physician. Finally, patient number was considered as the number of patients that a physician is caring for, which represents the physician's workload. All of the variables mentioned above are defined in Table 1. In addition, we used multiple linear regression to test the model in SPSS software (SPSS Inc, Chicago, IL, USA), and took the logarithmic value of the patient number, gift price, service price, and gift number to stabilize the variance. Our empirical models are summarized as follows:

(1) $\text{Phy\_Answer}=\alpha_{ij} + \alpha_1 \times Gift\_Price_{ij} + \alpha_2 \times Gift\_Number_{ij} + \alpha_3 \times Service\_Price_{ij} + \alpha_4 \times Control_j + \varepsilon_j$

(2) $\text{Phy\_Answer}=\alpha_{ij} + \alpha_1 \times Gift\_Price_{ij} + \alpha_2 \times Gift\_Number_{ij} + \alpha_3 \times Service\_Price_{ij} + \alpha_4 \times Gift\_Number_{ij} \times Gift\_Price_{ij} + \alpha_5 \times Service\_Price_{ij} \times Gift\_Price_{ij} + \alpha_6 \times Control_j + \varepsilon_j$

where i denotes a dialog and j denotes a physician. α_{ij} are the coefficients to be estimated. $Control_j$ represents the control variables for physician j, and ε_j is the standard error.

Table 1. Descriptions of model variables.

Variables and symbols	Variable types
Control variables	
Phy_Style	Counting variable
Ques_comp	Counting variable
Title	Binary variable (Title is chief physician: 1, otherwise: 0)
Hospital_Level	Binary variable (Hospital level is A: 1, otherwise: 0)
Recommendation	Counting variable
Patient_number	Counting variable
Independent variables	
Gift_Price	Counting variable
Service_Price	Counting variable
Gift_Number	Counting variable
Dependent variable	
Phy_Answer	Counting variable

Results

Descriptive Statistics and Correlations

Descriptive statistics and correlations for the key variables used in the analysis are presented in Multimedia Appendix 1. The results showed a correlation between gift price and the quality of physicians' online consultation service. In addition, correlations between the independent variable and control variables were relatively weak, which helped to yield more stable results.

Empirical Results

The empirical results of the model are summarized in Table 2, demonstrating support for all three of our hypotheses.

Model 1 shows the fitting degree of control variables and the dependent variable (adjusted R^2=0.04, F Change=17.003,

P<.001). In Model 2 (adjusted R^2=0.051, *F* Change=14.907, *P*<.001), gift price had a significant positive impact on the quality of physicians' online consultation services, which supports hypothesis 1. In model 3 (R^2=0.059, *F* change=15.079, *P*<.001), we added the interaction term of gift number and gift price based on model 2, showing that gift number plays a negative role in regulating the relationship between gift price and the quality of physicians' online consultation service, which supports hypothesis 2. Similarly, the results of model 4 (adjusted R^2=0.063, *F* change=16.613, *P*<.001) showed that service price negatively moderates the relationship between gift price and the quality of physicians' online consultation service, supporting hypothesis 3.

Table 2. Model test results.

Variables	Model 1		Model 2		Model 3		Model 4	
	β (SE)	*P* value	β (SE)	*P* value	β (SE)	*P* value	β (SE)	*P* value
Intercept	117.974(26.243)	<.001	85.143 (27.589)	.002	19.946 (30.823)	.52	–16.145 (33.024)	.63
Phy_Style	1.119 (0.172)	<.001	1.075 (0.175)	<.001	1.067 (0.174)	<.001	1.087 (0.174)	<.001
Ques_comp	0.257 (0.057)	<.001	0.250 (0.058)	<.001	0.250 (0.057)	<.001	0.252 (0.057)	<.001
Title	–3.732 (2.945)	.21	–4.325 (2.946)	.14	–4.789 (2.935)	.10	–4.435 (2.928)	.13
Hospital_Level	–3.558 (3.680)	.33	–2.167 (3.721)	.56	–2.118 (3.705)	.57	–3.096 (3.702)	.40
Recommendation	–15.597 (7.536)	.04	–14.559 (7.511)	.05	–15.472 (7.480)	.04	–13.651 (7.466)	.07
Patient_Number	–5.166 (1.877)	.006	3.496 (3.124)	.26	4.181 (3.114)	.18	4.732 (3.113)	.13
Gift_Price	N/A[a]	N/A	4.941 (1.523)	.001	31.286 (5.852)	<.001	42.321 (6.960)	<.001
Service_Price	N/A	N/A	–4.852 (1.812)	.007	–5.184 (1.805)	.004	18.870 (4.672)	<.001
Gift_Number	N/A	N/A	–5.282 (2.165)	.02	7.407 (3.472)	.03	–6.487 (2.162)	.003
Gift_Number*Gift_Price	N/A	N/A	N/A	N/A	–5.080 (1.090)	<.001	N/A	N/A
Service_Price*Gift_Price	N/A	N/A	N/A	N/A	N/A	N/A	–9.245 (1.680)	<.001

[a]N/A: not applicable; the variable was not included in the model.

Robustness Check

The quality of the information support paragraphs, instead of the complete answer, was used to test the robustness of our model. Specifically, we used the average word count of a physician's answer to measure the quality of the physician's online consultation service, but ignored the content included in the answer. However, online consultation services provided by physicians generally include two parts: information support and emotional support [46], which are the main functions in online health communities [47]. Therefore, we divided every answer of physicians into these two parts by manual annotation and explored the impacts of gift-giving behaviors on these components separately.

First, an expert was invited to define information support and emotional support according to related knowledge and real data. Information support was defined as the provision and exchange of information related to medical technology and medical process, such as diseases (diagnosis, prescription, treatment, and notes during treatment), hospitalization, registration, and others. In other words, information support refers to physicians answering professional questions and providing information. Emotional support was defined as expressions that are clearly not related to medical technology from physicians, such as gratitude, sympathy, comfort, support, encouragement, respect, courtesy, and responsibility. Care, supervision, active greetings, and offering active solutions to problems from physicians are all examples of emotional support; other cases should be judged flexibly according to the context. We then selected two graduate students with professional backgrounds to annotate 18,392 physician answers manually based on these definitions. The consistency check showed that the kappa value of information support was 0.85, whereas that of emotional support was 0.78. Finally, inconsistent data were annotated again by the expert. Descriptive statistics and correlations for all variables are shown in Multimedia Appendix 2. These results reflect that physicians typically pay more attention to patients' information needs (mean 38.870) than to their emotional needs (mean 3.595). Therefore, we decided to use the information support of physicians to check the robustness of our model, and empirical results are presented in Table 3.

We found a significant positive impact between gift price and the information support of physicians (Model 1: R^2=0.025, *F* Change=11.086, *P*<.001; Model 2: R^2=0.033, *F* Change=9.767, *P*<.001; Model 3: R^2=0.035, *F* change=9.457, *P*<.001; Model 4: R^2=0.038, *F* change=10.297, *P*<.001). Moreover, both gift number and service price negatively moderated the impact of gift price on physicians' information support. Hence, these three results are consistent with the main model, indicating that our model is robust. We also found that gift price positively affected the emotional support of physicians, but there was no moderating effect of gift number and service price (Model 5: R^2=0.071, *F* change=20.913, *P*<.001; Model 6: R^2=0.071, *F* change=18.298, *P*<.001).

Table 3. Model robustness test.

Variables	Model 1[a]		Model 2[a]		Model 3[a]		Model 4[a]		Model 5[b]		Model 6[b]	
	β (SE)	*P* value	β (SE)	*P* value	β (SE)	*P* value	β (SE)	*P* value	β (SE)	*P* value	β (SE)	*P* value
Intercept	68.062 (22.366)	.002	46.429 (23.563)	.05	15.961 (26.411)	.55	–13.749 (28.299)	.63	18.724 (3.320)	<.001	16.429 (3.999)	<.001
Phy_Style	0.767 (0.146)	<.001	0.741 (0.150)	<.001	0.737 (0.149)	<.001	0.748 (0.149)	<.001	0.017 (0.021)	.42	0.017 (0.021)	.42
Ques_comp	0.207 (0.049)	<.001	0.205 (0.049)	<.001	0.205 (0.049)	<.001	0.207 (0.049)	<.001	0.021 (0.007)	.003	0.021 (0.007)	.003
Title	–1.896 (2.510)	.45	–2.243 (2.516)	.37	–2.46 (2.515)	.33	–2.309 (2.509)	.36	–0.098 (0.355)	.78	–0.101 (0.355)	.78
Hospital_Level	1.230 (3.136)	.70	2.390 (3.178)	.45	2.413 (3.174)	.45	1.838 (3.172)	.56	–3.344 (0.448)	<.001	–3.365 (0.448)	<.001
Recommendation	–7.621 (6.423)	.24	–6.587 (6.415)	.31	–7.014 (6.409)	.27	–6.048 (6.398)	.35	–2.377 (0.904)	.009	–2.356 (0.904)	.009
Patient_Number	–3.328 (1.599)	.04	2.857 (2.668)	.28	3.177 (2.668)	.23	3.591 (2.667)	.18	–0.923 (0.376)	.01	–0.895 (0.377)	.02
Gift_Price	N/A[c]	N/A	2.720 (1.301)	.04	15.032 (5.014)	.003	24.929 (5.964)	<.001	0.939 (0.183)	<.001	1.786 (0.843)	.03
Service_Price	N/A	N/A	–3.954 (1.547)	.01	–4.110 (1.547)	.008	10.139 (4.004)	.01	0.584 (0.218)	.007	1.122 (0.566)	.05
Gift_Number	N/A	N/A	–3.675 (1.849)	.05	2.254 (2.975)	.45	–4.391 (1.853)	.02	–0.322 (0.260)	.22	–0.349 (0.262)	.18
Gift_NumberGift_Price	N/A	N/A	N/A	N/A	–2.374 (0.934)	.01	N/A	N/A	N/A	N/A	N/A	N/A
Service_PriceGift_Price	N/A	N/A	N/A	N/A	N/A	N/A	–5.493 (1.440)	<.001	N/A	N/A	–0.210 (0.203)	.30

[a]Information support.

[b]Emotional support.

[c]N/A: not applicable; the variable was not included in the model.

Discussion

Principal Findings

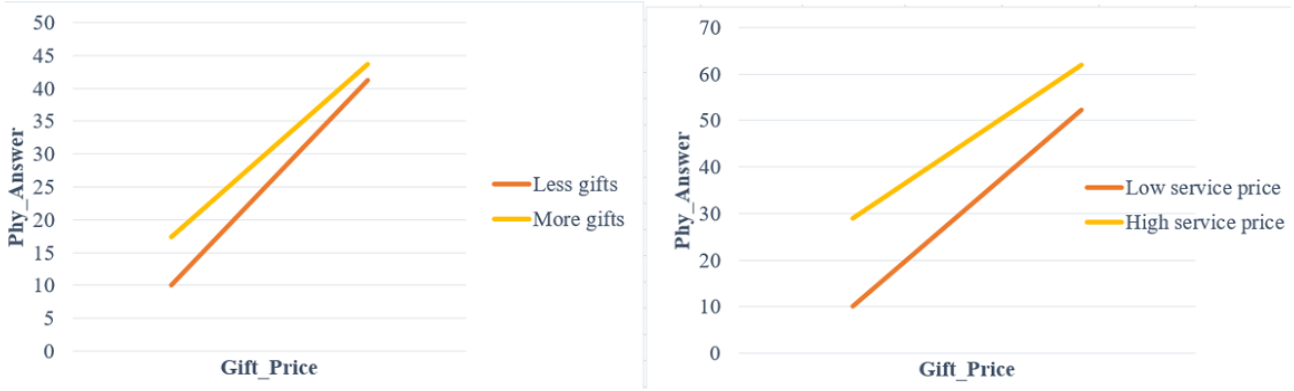
Our empirical results supported all three of our hypotheses. First, in online medical consultations, gift price positively affects the quality of physicians' online consultation service. This suggests that in response to patients' mental and economic efforts, physicians will reciprocate with better online service, which contributes to establishing more stable physician-patient relationships. This result confirms the existence of the principles of reciprocity and equality in online physician-patient interactions. Furthermore, these principles are also widespread in other online communities. For example, researchers have verified that an increase in the number of reciprocity messages the actor broadcasts in online social networks increases the reciprocity reactions from his or her audience [21]. Similarly, some scholars have found that there is a consistent reciprocal mode between the information users publish and the answer they receive in an online gaming community [48].

Second, the number of online gifts physicians receive plays a negative role in regulating the impact of gift price on the quality of physicians' online consultation service, proving that a marginal diminishing effect exists not only in the field of economics but also in people's productivity, life, and social management. Some researchers have studied the relationship between individual income and happiness based on this theory, proving that higher income increases happiness in developing countries, whereas this effect is minimal in developed countries [11].

Third, service price negatively moderates the relationship between gift price and the quality of physicians' online consultation service. The gift price in Good Physician Online ranges from 1 to 200 yuan, accounting for about one quarter of the service price on average. In addition, the service price could be used as the comparison level for physicians when they judge the value of gifts because patients always pay for the consultation before sending gifts. Comparison level theory was originally developed to study people's assessment of interpersonal relationships, such as the perception of marital relationships [49]. Subsequently, scholars have applied this theory in many different contexts. For example, researchers confirmed that workers' reported satisfaction levels were inversely related to their comparison wage rates [50] and idealized advertising images reduced women's satisfaction with their attractiveness [51]. Collectively, these studies demonstrated that comparison level theory is a psychological phenomenon

that widely exists in several aspects of human society. For further explanation of these interaction effects, see Figure 3.

Figure 3. Interaction effects of gift number and service price.



Finally, both information support and emotional support could be enhanced by physicians when they receive gifts with higher prices. However, compared with patients' emotional needs, physicians in online health communities tend to pay more attention to their information needs.

Theoretical Implications

Our study contributes new knowledge in several key ways. First, despite previous research about the impact of online gifts in social networks and live broadcast platforms [33,52-55], few scholars have paid attention to the gifts provided in online health communities. Furthermore, research related to gift-giving behaviors in an online health community verified that online gifts from patients could improve the speed of physicians' answers during the consultation [8], but ignored the potential effect of gift price. Our study is among the first to use real data to empirically examine the effect of gift price in an online health community, which is a universally beneficial sector.

Second, SET was proposed in the 1960s and has since been widely applied in various fields. In recent years, SET has often been used to explain user behaviors in online contexts [22,56-58]. Our study further adds to this literature by verifying that the physician-patient interaction is also a form of social exchange.

Third, scholars have long highlighted the importance of emotional support [59], but few studies have empirically investigated this assumption. Our study revealed the actual situation of information and emotional support in the online medical field using a manual annotation approach. In addition, the text of physician-patient interactions contains an abundance of valuable information, which has inspired us to adopt a text mining analysis in future research.

Practical Implications

The original intention of online gifts in online health communities is to provide patients with a channel to express gratitude to physicians. However, this channel has been questioned since its launch from an ethics perspective because physicians can receive economic benefits from this practice. Although patients purchase gifts voluntarily, they may still feel embarrassed and uncomfortable [5]. Moreover, some physicians have recommended refusing gifts from patients under the pressure of public opinion and professional ethics [60,61]. All of these questions motivated our reflections about the validity of online gifts in online health communities.

The results of our study indicate that gift price has a significant positive impact on the quality of physicians' online consultation service, and the impact is stronger for physicians that receive fewer gifts or a lower service price. These conclusions could provide practical guidance for patients, but such speculation is beyond the scope of this study. Furthermore, we do not encourage sending gifts with higher prices to physicians as this may lead to unhealthy gift-giving trends in online health communities and undermine online physician-patient relationships. Thereby, we propose some suggestions from the perspective of platform management. First, compared with service price, the gift price in some online health communities is sufficiently high to affect the interactions between physicians and patients; thus, we recommended lowering the gift price or narrowing the price range to mitigate this impact. Second, we suggest that more value should be placed on the function of emotional expression. Specifically, patients could be granted more permissions such as sending images or voice messages when they write their greetings associated with gifts. Third, in addition to online gifts, other free channels for emotional expression could be provided to relieve the discomfort of patients.

Finally, the results of our robustness check revealed that physicians pay more attention to patients' information needs rather to their emotional needs when they provide online consultation services. However, emotional support is more effective than information support in alleviating patients' poor conditions in online health communities [62]. Therefore, we recommend that physicians could attach more importance to patients' emotional needs during online consultations, which could contribute to achieving a better treatment effect and improved patient satisfaction.

Limitations and Future Research

Several limitations and prospects of this study must be considered. First, we focused on only one context, which helps us improve the internal validity but may also reduce the generalizability of our findings. Therefore, future research should be performed to validate our results in other contexts. Second, there is no available mature criterion to measure online medical service quality. In this study, we used both the word

count and content of physicians' answers to measure service quality. However, it is still difficult to fully represent the actual quality of physicians' services. In the future, we will continue to explore new ways to measure the quality of physicians' online medical services. Third, the possible cumulative effect of gifts was ignored in our study, which could also be explored in future research. Fourth, multiple methods should be used to understand the true significance of these key findings, such as a quality study.

Conclusions

Despite the original intention of online gifts as offering a new channel for patients to thank physicians, people remain suspicious of the validity of online gifts. Our study offers a better understanding of the impact of online gifts and contributes to settling existing disputes in this field. We found that the quality of physicians' online consultation service could be affected by the price of gifts received from patients, which implies that online gifts are more than a simple channel for patients to express their emotions. However, we hold that it is important to consider online gifts from an objective and rational perspective since online gifts in online health communities are still new. We believe that this paper can help provoke new ideas and perspectives about the validity of online gifts in online health communities.

Acknowledgments

This study was funded by the National Natural Science Foundation of China (91646104).

Authors' Contributions

All authors contributed to this paper. YW, CX, HW, and NL conceived and designed the study, developed the research model, conducted data collection and analysis, and drafted as well as modified the manuscript. All authors approved the final version of the manuscript for submission

Conflicts of Interest

None declared.

Multimedia Appendix 1
Descriptive statistics and correlations.
[DOCX File , 14 KB - jmir_v22i5e15685_app1.docx]

Multimedia Appendix 2
Descriptive statistics and correlations in robustness check.
[DOCX File , 14 KB - jmir_v22i5e15685_app2.docx]

References

1. Joy A. Gift Giving in Hong Kong and the Continuum of Social Ties. J Consum Res 2001 Sep 01;28(2):239-256. [doi: 10.1086/322900]
2. Ruth J, Otnes C, Brunel F. Gift Receipt and the Reformulation of Interpersonal Relationships. J Consum Res 1999 Mar;25(4):385-402. [doi: 10.1086/209546]
3. Chiu Y, Smith KC, Morlock L, Wissow L. Gifts, bribes and solicitions: print media and the social construction of informal payments to doctors in Taiwan. Soc Sci Med 2007 Feb;64(3):521-530. [doi: 10.1016/j.socscimed.2006.09.018] [Medline: 17070971]
4. Zhu W, Wang L, Yang C. Corruption or professional dignity: An ethical examination of the phenomenon of "red envelopes" (monetary gifts) in medical practice in China. Dev World Bioeth 2018 Mar 21;18(1):37-44. [doi: 10.1111/dewb.12152] [Medline: 28544334]
5. Zhang X. Iqilu News. 2016 Mar 14. Physicians in Good Physician Online Website were exposed at accepting gifts! WeChat pays attention to red envelopes in disguise URL: http://news.iqilu.com/shandong/kejiaoshehui/20160314/2717916.shtml [accessed 2019-06-15]
6. Yan H. Sohu. 2016 Mar 22. Are online virtual gifts to physicians URL: https://www.sohu.com/a/64868755_161794 [accessed 2019-06-15]
7. Song S. Complaints about Good Physician Online website collecting red envelopes, and the truth is that. 2016 Mar 16. URL: https://www.sohu.com/a/63908063_377326 [accessed 2019-06-15]
8. Zhao W, Guo X, Wu T, Geng J. The Impacts of Patients' Gift Giving Behavior on Physicians' Service Quality. 2017 Oct 28 Presented at: International Conference on Smart Health 2017; 2017; Hong Kong, China. [doi: 10.1007/978-3-319-67964-8_5]
9. Farmer M, Heath A. Rational Choice and Social Exchange. Br J Sociol 1977 Jun 1;28(2):254-255 [FREE Full text] [doi: 10.2307/590216]

10. Homans GC. Social Behavior as Exchange. Am J Sociol 1958 May;63(6):597-606. [doi: 10.1086/222355]
11. Easterlin RA. Diminishing Marginal Utility of Income? Caveat Emptor. Soc Indic Res 2005 Feb;70(3):243-255. [doi: 10.1007/s11205-004-8393-4]
12. Cropanzano R, Mitchell MS. Social Exchange Theory: An Interdisciplinary Review. J Manag 2016 Jul;31(6):874-900. [doi: 10.1177/0149206305279602]
13. Malinowski B. Argonauts of the Western Pacific: An account of native enterprise and adventure in the archipelagoes of Melansian New Guinea. London: Routledge; Sep 10, 2002:1-616.
14. Mauss M, Halls W. The Gift: The Form and Reason for Exchange in Archaic Societies. New York: WW Norton and Company; 2000.
15. Firth R. Themes In Economic Anthropology. London: Routledge; 2020.
16. Sahlins M. Stone Age Economics. London: Routledge; Jan 1, 1972.
17. Gouldner AW. The Norm of Reciprocity: A Preliminary Statement. Am Sociol Rev 1960 Apr;25(2):161-178. [doi: 10.2307/2092623]
18. de Montlibert C, Thibaut JW, Kelley HH. The Social Psychology of Groups. Revue Française de Sociologie 1960 Jul;1(3):374-186. [doi: 10.2307/3319917]
19. Bierstedt R, Blau PM. Exchange and Power in Social Life. Am Sociol Rev 1965 Oct;30(5):789-790. [doi: 10.2307/2091154]
20. Emerson RM. Power-Dependence Relations. Am Sociol Rev 1962 Feb;27(1):31-41. [doi: 10.2307/2089716]
21. Surma J. Social exchange in online social networks. The reciprocity phenomenon on Facebook. Comput Commun 2016 Jan;73:342-346. [doi: 10.1016/j.comcom.2015.06.017]
22. Shiau W, Luo MM. Factors affecting online group buying intention and satisfaction: A social exchange theory perspective. Comput Hum Behav 2012 Nov;28(6):2431-2444. [doi: 10.1016/j.chb.2012.07.030]
23. Chou S, Hsu C. Understanding online repurchase intention: social exchange theory and shopping habit. Inf Syst E-Bus Manage 2015 Jan 9;14(1):19-45. [doi: 10.1007/s10257-015-0272-9]
24. Gefen D, Karahanna E, Straub DW. Trust and TAM in Online Shopping: An Integrated Model. MIS Quart 2003 Mar;27(1):51-90. [doi: 10.2307/30036519]
25. Luo X. Trust production and privacy concerns on the Internet. Indust Market Manage 2002 Feb;31(2):111-118. [doi: 10.1016/s0019-8501(01)00182-1]
26. Tong Y, Wang X, Teo H. Understanding the Intention of Information Contribution to Online Feedback Systems from Social Exchange and Motivation Crowding Perspectives. 2007 Presented at: 40th Hawaii International International Conference on Systems Science (HICSS-40 2007); January 3-6, 2007; Hawaii. [doi: 10.1109/HICSS.2007.585]
27. Liu Z, Min Q, Zhai Q, Smyth R. Self-disclosure in Chinese micro-blogging: A social exchange theory perspective. Inform Manage 2016 Jan;53(1):53-63. [doi: 10.1016/j.im.2015.08.006]
28. Tsai JC, Kang T. Reciprocal intention in knowledge seeking: Examining social exchange theory in an online professional community. Int J Inf Manage 2019 Oct;48:161-174. [doi: 10.1016/j.ijinfomgt.2019.02.008]
29. Jin B, Park JY, Kim H. What makes online community members commit? A social exchange perspective. Behav Inf Technol 2010 Nov;29(6):587-599. [doi: 10.1080/0144929x.2010.497563]
30. Yan Z, Wang T, Chen Y, Zhang H. Knowledge sharing in online health communities: A social exchange theory perspective. Inf Manage 2016 Jul;53(5):643-653. [doi: 10.1016/j.im.2016.02.001]
31. Zhao T, Du R. Why People Are Willing to Provide Social Support in Online Health Communities: Evidence from Social Exchange Perspective. 2018 Oct 26 Presented at: International Conference on Smart Health 2018; 2018; Wuhan. [doi: 10.1007/978-3-030-03649-2_12]
32. Guo S, Guo X, Fang Y, Vogel D. How Doctors Gain Social and Economic Returns in Online Health-Care Communities: A Professional Capital Perspective. J Manage Inf Syst 2017 Aug 17;34(2):487-519. [doi: 10.1080/07421222.2017.1334480]
33. Kim H, Kankanhalli A, Lee S. Examining Gifting Through Social Network Services: A Social Exchange Theory Perspective. Inf Syst Res 2018 Dec;29(4):805-828. [doi: 10.1287/isre.2017.0737]
34. Pervan SJ, Bove LL, Johnson LW. Reciprocity as a key stabilizing norm of interpersonal marketing relationships: Scale development and validation. Ind Market Manage 2009 Jan;38(1):60-70. [doi: 10.1016/j.indmarman.2007.11.001]
35. Malinowski B. Crime and custom in savage society. Nature 1927 Apr 1;119(3000):633-633. [doi: 10.1038/119633c0]
36. Sutton S. Social-psychological approaches to understanding addictive behaviours: attitude-behaviour and decision-making models. Br J Addict 1987 Apr;82(4):355-370. [doi: 10.1111/j.1360-0443.1987.tb01492.x] [Medline: 3472583]
37. Rachlin H. Diminishing marginal value as delay discounting. J Exp Anal Behav 1992 May;57(3):407-415 [FREE Full text] [doi: 10.1901/jeab.1992.57-407] [Medline: 1602271]
38. Dawson G. Under what circumstances might demand and price rise? Does such an observation contradict the law of diminishing marginal utility? Econ Affairs 1989 Oct;10(1):41-42. [doi: 10.1111/j.1468-0270.1989.tb00709.x]
39. About us- Good Physician Online. URL: https://www.haodf.com/helper/present.htm [accessed 2019-06-20]
40. Yi Kahui. IDF Global Diabetes Map (ninth edition) have released in 2019!. 2019 Nov 15. URL: https://www.sohu.com/a/354003012_489312 [accessed 2020-04-25]
41. Chen H, Yu P, Hailey D, Wang N. Methods for assessing the quality of data in public health information systems: a critical review. Stud Health Technol Inform 2014;204(5):13-18. [Medline: 25087521]

42. Whitten P, Nazione S, Lauckner C. Tools for assessing the quality and accessibility of online health information: initial testing among breast cancer websites. Inform Health Soc Care 2013 Dec 19;38(4):366-381. [doi: 10.3109/17538157.2013.812644] [Medline: 23957628]
43. Weiner SS, Horbacewicz J, Rasberry L, Bensinger-Brody Y. Improving the Quality of Consumer Health Information on Wikipedia: Case Series. J Med Internet Res 2019 Mar 18;21(3):e12450 [FREE Full text] [doi: 10.2196/12450] [Medline: 30882357]
44. Chen M, Li Y, Liu X. Assessment of online reference performance in gastrointestinal diseases. Chinese J Med Library Inform Sci 2016;25(11):66-69. [doi: 10.3969/j]
45. Mudambi SM, Schuff D. Research Note: What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com. MIS Quart 2010;34(1):185-200. [doi: 10.2307/20721420]
46. Stewart KA, Segars AH. An Empirical Examination of the Concern for Information Privacy Instrument. Inform Syst Res 2002 Mar;13(1):36-49. [doi: 10.1287/isre.13.1.36.97]
47. Nambisan P. Information seeking and social support in online health communities: impact on patients' perceived empathy. J Am Med Inform Assoc 2011 May 01;18(3):298-304 [FREE Full text] [doi: 10.1136/amiajnl-2010-000058] [Medline: 21486888]
48. Pan Y, Kuo L, Lee J. Sociability Design Guidelines for the Online Gaming Community: Role Play and Reciprocity. 2007 Presented at: International Conference on Online Communities and Social Computing; July 22-27, 2007; Berlin p. 426-434. [doi: 10.1007/978-3-540-73257-0_47]
49. Sabatelli RM. The Marital Comparison Level Index: A Measure for Assessing Outcomes Relative to Expectations. J Marriage Family 1984 Aug;46(3):651-662. [doi: 10.2307/352606]
50. Clark AE, Oswald AJ. Satisfaction and comparison income. J Public Econ 1996 Sep;61(3):359-381. [doi: 10.1016/0047-2727(95)01564-7]
51. Richins ML. Social Comparison and the Idealized Images of Advertising. J Consum Res 1991 Jun;18(1):71. [doi: 10.1086/209242]
52. Strelnikova A, Polukhina E. Exploring the Russian Online Gift-Exchange Communities: The Results of Nethnographic Approach. Qual Rep 2015 Dec;20(12):2041-2049.
53. Shmargad Y, Watts JK. When Online Visibility Deters Social Interaction: The Case of Digital Gifts. J Interact Market 2016 Nov;36:1-14. [doi: 10.1016/j.intmar.2016.01.004]
54. Lampel J, Bhalla A. The Role of Status Seeking in Online Communities: Giving the Gift of Experience. J Comput-Mediat Commun 2007;12(2):434-455. [doi: 10.1111/j.1083-6101.2007.00332.x]
55. Yu E, Jung C, Kim H, Jung J. Impact of viewer engagement on gift-giving in live video streaming. Telemat Informat 2018 Aug;35(5):1450-1460. [doi: 10.1016/j.tele.2018.03.014]
56. Wang Y, Xiang D, Yang Z, Ma S. Unraveling customer sustainable consumption behaviors in sharing economy: A socio-economic approach based on social exchange theory. J Clean Product 2019 Jan;208:869-879. [doi: 10.1016/j.jclepro.2018.10.139]
57. Boateng H, Kosiba JPB, Okoe AF. Determinants of consumers' participation in the sharing economy. Int J Contemp Hospitality Mngt 2019 Feb 11;31(2):718-733. [doi: 10.1108/IJCHM-11-2017-0731]
58. Hsu CS, Chou SW, Min HT. Understanding Knowledge Management Behavior From A Social Exchange Perspective. 2017 Jul Presented at: 5th International Conference on Research and Innovation in Information Systems (ICRIIS2017); 2017; Langkawi Island, Malaysia. [doi: 10.1109/ICRIIS.2017.8002492]
59. Clark PA, Drain M, Malone MP. Addressing Patients' Emotional and Spiritual Needs. Joint Commiss J Qual Safety 2003 Dec;29(12):659-670. [doi: 10.1016/s1549-3741(03)29078-x]
60. Sun Z. The statement of thoughtful gifts!!! (Original). 2015 Nov 28. URL: https://www.haodf.com/zhuanjiaguandian/sunzhanxue_3848764639.htm [accessed 2019-06-15]
61. Chen W. About online gifts in Good physician online website. 2018 Mar 4. URL: https://m.haodf.com/touch/zhuanjiaguandian/haodf_5887966498.htm [accessed 2019-06-15]
62. Yan L, Tan Y. Feeling Blue? Go Online: An Empirical Study of Social Support Among Patients. Inform Syst Res 2014 Dec;25(4):690-709. [doi: 10.1287/isre.2014.0538]

Abbreviations

SET: social exchange theory

Edited by G Eysenbach; submitted 29.07.19; peer-reviewed by Q Ye, Y Song; comments to author 29.09.19; revised version received 15.10.19; accepted 21.12.19; published 05.05.20.

Please cite as:
Wang Y, Wu H, Xia C, Lu N
Impact of the Price of Gifts From Patients on Physicians' Service Quality in Online Consultations: Empirical Study Based on Social Exchange Theory
J Med Internet Res 2020;22(5):e15685
URL: http://www.jmir.org/2020/5/e15685/
doi:10.2196/15685
PMID:32369028

Original Paper

The Impact of Video-Mediated Communication on Closed Wound Assessments in Postoperative Consultations: Conversation Analytical Study

Wyke J P Stommel[1], PhD; Harry van Goor[2], PhD; Martijn W J Stommel[2], PhD

[1]Centre for Language Studies, Radboud University, Nijmegen, Netherlands

[2]Department of Surgery, Radboud University Medical Center, Nijmegen, Netherlands

Corresponding Author:
Wyke J P Stommel, PhD
Centre for Language Studies
Radboud University
Erasmusplein 1
Nijmegen, 6500 HD
Netherlands
Phone: 31 243612482
Email: w.stommel@let.ru.nl

Abstract

Background: Research on the use of video-mediated technology for medical consultations is increasing rapidly. Most research in this area is based on questionnaires and focuses on long-term conditions. The few studies that have focused on physical examinations in video consultations indicated that it poses challenges for the participants. The specific activity of wound assessment through video in postsurgery consultations has not yet been studied. Furthermore, a comparative analysis of face-to-face and video settings on the moment-to-moment organization of such an activity is original.

Objective: The aim of this study was to examine the impact of video technology on the procedure of postsurgery wound assessment and its limits.

Methods: We recorded 22 postoperative video consultations and 17 postoperative face-to-face consultations. The primary purpose of the consultation was to inform the patient about the final pathology results of the resected specimen, and the secondary purpose was to check on the patient's recovery, including an assessment of the closed wound. The recordings were transcribed in detail and analyzed using methods of conversation analysis.

Results: The way that an assessment of the wound is established in video consultations differs from the procedure in face-to-face consultations. In the consultation room, wound assessments overwhelmingly (n=15/17) involve wound showings in the context of surgeons reporting their observations formatted with evidentials ("looks neat") and subsequently assessing what these observations imply or what could be concluded from them. In contrast, wound assessments in video consultations do not tend to involve showing the wound (n=3/22) and, given the technological restrictions, do not involve palpation. Rather, the surgeon invites the patient to assess the wound, which opens up a sequence of patient and physician assessments where diagnostic criteria such as redness or swollenness are made explicit. In contrast to observations in regular consultations, these assessments are characterized by epistemic markers of uncertainty ("I think," "sounds...good") and evidentials are absent. Even in cases of a potential wound problem, the surgeon may rely on questioning the patient rather than requesting a showing.

Conclusions: The impact of video technology on postoperative consultations is that a conclusive wound assessment is arrived at in a different way when compared to face-to-face consultations. In video consultations, physicians enquire and patients provide their own observations, which serve as the basis for the assessment. This means that, in video consultations, patients have a fundamentally different role. These talking-based assessments are effective unless, in cases of a potential problem, patient answers seem insufficient and a showing might be beneficial.

(J Med Internet Res 2020;22(5):e17791) doi:10.2196/17791

KEYWORDS

video consultation; remote consultation; physical examination; assessment; patient-physician relationship; conversation analysis

Introduction

Video consultations are generally found promising for use in the medical domain, especially due to advantages such as remoteness, convenience for patients and informal caregivers, and reduced anxiety [1-4]. However, the implementation of video consultations into real-world settings is complex. Most research in this area focuses on long-term conditions and is based on questionnaires to elicit patient and clinician experiences, reporting both positive and negative experiences with video consultations. The experiences often seem to depend on the context (eg, a long-term condition in which the clinician and patient have a pre-existing relationship, and on whether both parties are confident in dealing with technical issues [5,6]). Adaptation to the context can be accomplished by involving the patient in the choice of consultation modality. In a comparative study on video versus face-to-face consultations in follow-up care after colorectal cancer surgery, video consultations based on patient preference were shown to be equivalent to face-to-face consultations in terms of patient satisfaction and perceived quality of care [7]. The type of patients that might be most suitable for video consultations is unclear, but it is recognized that patients' reasons and ability to use video consultation may change over time and with experience [8]. One of the advantages of the video format is that it affords visual access, which at least in theory enables physicians to visually assess what patients show. Nevertheless, physical examination has been regarded as problematic in the video setting [9]. Patients' self-examinations in front of the camera, such as measuring weight, blood pressure, heart rate and rhythm, and oxygen saturation, appeared to be challenging in various respects [10]. One of the challenges was that patients had to do a physical examination while simultaneously making it visible to the clinician. Hence, visual access may not be just an advantage; it can also create new problems.

Conversation analytical studies of medical video consultations are beginning to uncover microlevel dimensions and challenges of video-mediated consultations [10-12], sometimes explicitly in comparison with copresent consultations [13]. Pappas and Seale [11,12] analyzed medical video consultations with a primary care physician or nurse and a patient at one end of the connection and a consultant (a medical specialist) at the other. The professional who was with the patient and, therefore, had direct perceptual access to the patient's body, assessed the patient's foot and used the visual channel to demonstrate the assessment to the physician on the other end [12]. Seuren et al [10] identified various challenges of video consultations in secondary care related to instructions for patients to self-measure oxygen levels and manipulate the camera and the body to capture what should be viewed by the physician.

A key domain of interest to medical video consultations is the physical examination, which requires the physician's visual access to the patient's body. Visual access is an affordance [14] of video-mediated interactions, despite the "fractured ecologies" [15] of the patient and the physician inherent to the interaction. It has been found that, by doing physical examinations, the remote physician transposes observational authority to the patient's site [12]. Relatedly, examination conducted by patients themselves may enhance their autonomy with regard to their own health [10]. Hence, physical examination in video consultations may have the potential to instigate a shift in the physician-patient relationship or, more broadly, in the way that medicine is practiced.

When physicians examine patients, they may communicate the findings of their observations to the patient [16-18]. Simultaneously with the act of examination, physicians may produce talk that is subordinated to the examination, which is called online commentary. There is usually no mutual gaze and no response from the patient due to a lack of shared access to the object of evaluation (eg, a physician inspecting a patient's ear). Alternatively, patients may be invited to provide an initial self-assessment as long as they have access to the object of examination.

There are two primary formats for communication along with physical examination, namely, reports of observations and assessments of what is observed [16,19]. In cases of an *observation report* (eg, "I don't see"), the conclusions such as "looks good" should be drawn by the patient. With *assessments* of what is observed (eg, "that looks good"), it is the physician who presents a conclusion. Essentially, the power of both formats lies in the physician's epistemic "ecological advantage" [16] to be able to perceptually (seeing, hearing, feeling) assess the state of the patient's body. The criteria or "codes" for the evaluation that are discursively constructed in the interaction serve as an apparatus of the physician's professional vision [20]. The ecological advantage, thus, encompasses rights with regard to both the examination and constructing observation in certain assessment categories.

Assessments are evaluations of objects and events in talk-in-interaction [21]. Assessments can be elicited both by verbal actions (eg, questions, prior assessments) and by embodied conduct or experience (cf [21,22]). Goodwin and Goodwin [23] discern assessments on distinct levels of organization, with assessment activities as one such level. Assessment as an activity refers to multiple participants jointly producing an assessment in multiple turns, using intonation, overlap, intensifiers, nods, and other resources. Relevant to such assessment activities is that the participants have differential access to the assessable, which is reflected in their talk. For instance, saying "that sounds good" attends to the fact that the assessable was available through a coparticipant's description [23]. Displaying agreement on or producing concurring assessments is important in assessment activities. The same speaker can repeat an assessment; although, subsequent assessments may display diminished participation and, thus, bring the activity to a close. Overall, an activity of assessment is a structure that participants collaboratively bring to a climax and then withdraw from.

A specific occasion for the occurrence of assessments are "informative showings" [24], which involves showing something "new" such as the current state of the wound and a recipient who is informed by the showing. In the medical context, a showing enables the physician's professional vision [20] as a basis for assessment rather than that joint visual perception that is achieved by the showing (cf [25-27]).

The question this paper addresses is how assessments of a surgical closed wound are collaboratively produced in video consultations where the physician lacks direct perceptual access to the assessable, which is available in the face-to-face setting. This question provides insights to the ways that video technology as a mode of communication affects clinical practice.

Methods

The data consist of 39 video recordings of follow-up consultations after abdominal cancer surgery, including 17 copresent consultations (average length 13 minutes and 40 seconds) and 22 video consultations (average length 12 minutes and 20 seconds). The data were collected in the context of a study comparing the conversational organization of video-mediated consultations with regular consultations at the outpatient clinic during the first postoperative consultation after discharge [13]. The first postoperative consultation was chosen because of the potentially considerable burden of a visit to the outpatient clinic, and thus, video consultations had a potential advantage in this phase. The inclusion criteria were patients ≥18 years of age who had received abdominal cancer surgery. The exclusion criteria were an inability to give informed consent and a lack of proficiency in Dutch. A total of 39 patients participated (21 female and 18 male) often accompanied by one or more family members and 3 male surgeons who were experienced in video consulting before the start of the study. The type and complexity of the surgery was comparable for all patients; although, some were laparoscopic operations, which involved three or four small incisional wounds for the trocars, rather than one large wound. At discharge, the patients were informed about the follow-up consultation scheduled approximately 2 weeks after the operation. The goal of the follow-up consultation was explained as discussing the final pathology results and checking on recovery. The results sometimes involved bad news but were mostly brought on as a confirmation of what was expected. The question about recovery, including the wound assessments, usually came as a second order of business for the consultation [13]. The patients were offered follow-up consultations through video or as a regular consultation at the hospital. After they chose either of these options, they were informed about the study and asked to participate. They all gave their consent; although, 1 patient requested at a later time for us to not use the recording or transcript in any scientific publication or presentation of the study.

A waiver for medical ethical approval was obtained from Radboud Medical Center Ethical Committee in June 2017. The data were collected in June-July 2017 and March-June 2018. Each consultation was recorded using two cameras, one directed mainly at the surgeon and one at the patient and those that accompanied them, either in the consultation room or on the surgeon's desktop computer. The recordings, thus, reflect the real-life circumstances of the surgeon, who does not have access to whatever the patient sees or hears such as delays, perturbations, or sequential mismatches (cf [28]). The particulars of the ecological setup at patients' homes may have influenced some practical choices, for instance readjusting their body to the screen to make them visible rather than turning the camera toward the spot, which is another way to show things in video communication [29]. For the hospital recordings, the surgeons sometimes turned away the camera or even turned it off during the physical examination for ethical reasons; these recordings were not excluded from the data set, although they were inevitably inapt for detailed analysis of the examination. The consultations were transcribed based on conversation analysis conventions [30,31] (see Multimedia Appendix 1), and all names were replaced by pseudonyms.

To juxtapose assessments in copresent and video-mediated consultations, we first identified all wound assessment activities in the data and whether it involved a showing or not. The next step was to examine each case microanalytically using multimodal conversation analysis [32,33]. These analyses were inherently comparative, resulting in an understanding of the methods used for doing wound assessments contingent on the medium. In the next section, we discuss five illustrative cases of wound assessments, two in the copresent setting and three in the video setting, that are representative of our findings.

Results

Wound Assessment

We found that in the copresent setting, wounds were generally assessed on the basis of a showing of the wound. On the contrary, in the video-mediated setting, showings were rare. Table 1 provides an overview of the occurrence of wound assessment, wound assessment including showing, and no wound assessment.

Table 1. Frequency of showing-based and talk-based wound assessment in copresent and video-mediated consultations.

Group	Showing-based wound assessment, n (%)	Talk-based wound assessment, n (%)	No wound assessment, n (%)
Copresent (n=17)	15 (88)	1 (6)	1 (6)
Video-mediated (n=22)	3 (17)	12 (51)	7 (32)
Total (N=39)	18 (46)	13 (33)	8 (21)

In the following section, we first analyzed the face-to-face default method and then the default video-mediated communication (VMC) method. We found no communicative differences between laparoscopic wounds and other wounds. The two assessment procedures were mostly initiated by the physician who enquires about how the wound is healing. We also discuss one video consultation where the wound assessment is initiated by a patient who reports a potential problem. This allows for an in-depth understanding of the intricacies of wound assessment through video.

The Face-to-Face Default Method: Showing-Based Wound Assessments

Showing-based wound assessments are characterized by a relatively stable structure. It is initiated by the physician asking whether the wound(s) are healing well. The patient's answer is a tentatively positive evaluation of the wound, upon which the physician expands the sequence with a request to show the wound or an invitation to undress behind the curtain for a physical examination. During the showing or examination, the physician produces observation reports using evidentials and evaluations ("looks neat," "looks uneventful"). The evaluations tend to be rather brief with general descriptors like "neat," "good," and "uneventful." The excerpt in Figure 1 is an example of a copresent showing-based wound assessment. The physician is enquiring about the patient's recovery, having asked about fever and illness (data not shown) and then about the wounds (line 1).

Figure 1. Excerpt 1. Regular consultation number 25 (video time: 5:50).

```
1  DOC: =zijn de wondjes goed genezen?
       =did the small wounds heal well?
2  PAT: volgens mij *wel.
   Doc:             *firm nod
       according to me yes.
       I think so.
3  DOC: oke: kunt u *ut- ut #^hemd even omhoogtrekken (.) dat ik kan kijken?
                         #figure 2
   Doc:             *points to Pat's belly and moves forward
   Pat:                  ^-----------------------------------------------
       okay: can you- briefly pull up the- the shirt so that I can have a look?
4      (2.0)^#
   Pat: -----^stands up and takes out shirt
5  DOC: ohja *>nou<^ dat ziet er keurig uit ^*(.) ja;
       oh yeah >well< that looks neat       (.) yeah;
   Doc:      *bends forward
   Pat:           ^------------------------^turns hips, showing belly
   Doc:                                     *leans back in chair
6  PAT: ^volgens mij wel heh,
       according to me it does huh,
       I think so huh,
   Pat: ^--------------------
7  DOC: jaha:: (.) zeker ja;
       yea::h (.) it certainly does yeah;
   Pat:  -------------------
8      *(0.5)
   Doc: *nods
   Pat:  ------
9  DOC: °ziet er rustig uit°.
        °looks uneventful.°
        °healing well.°
   Pat:  ----------------------
10 DOC: ja;*^
       yeah;
   Doc:    *looks at pc screen
   Pat: ----^puts shirt back into jeans
```

The showing-based wound assessment activity begins with the physician's yes/no-question (line 1), asking whether the small wounds healed well. The patient confirms with an epistemic downgrade ("according to me" line 2), thus, making an independent assessment by the surgeon a relevant next action. The request to show the wound "can you- briefly pull up the shirt" (line 3) displays a relatively high entitlement [34], thus, building on the structure of wound assessment as projected by the patient's previous turn. The patient complies immediately (see line 3 and Figure 2), aligning with the activity. As soon as the wound is visible, the physician reports that it "looks neat" and confirms the patient's initial evaluation ("yeah" line 5). Note that the physician uses an evidential ("looks") to present a conclusion rather than an observation report, later rephrased as "looks uneventful" (line 9), which is produced softly and, thus, displays diminished participation and an orientation to activity closing [23].

Figure 2. Screenshot for excerpt 1 (video time: 5:59).

Giving visual access in the showing-based assessment activity does not need to be requested explicitly, as it was in Figure 1. The activity structure in the copresent setting allows for more subtle collaborative orientation to the relevance of showing as can be seen in Figure 3.

Figure 3. Excerpt 2. Regular consultation number 36 (video time: 7:37).

```
1  PAT2: =dus:=
         =so:=
2  DOC:  =.hh* en hoe is de wond genezen meneer [van dijk;
         =.hh and how is the wound healing mister [van dijk;
   Doc:      *points briefly at Pat belly
3  PAT2:                                        [KEUrig.
                                                [neatly.
4  PAT:  nou die is °netjes°
         well that looks °neat°
5  DOC:  ^ja?
          yeah?
   Pat:  ^#---
          #figure 4
6  PAT2: vinden wij wEl,
          we do think so,
   Pat:  ---------------
7  PAT3: ja das echt mooi.
         yeah that's really pretty.
   Pat:  -----------------^#stands up and pulls up shirt
8  DOC:  Oh ja (.) ja dat# ziet er netjes uit.
         Oh yeah (.) yeah that looks neat.
                        #figure 5
9  PAT2: keurig (heh),*
          neat (huh),
   Doc:               *nods
10 DOC:  °hartstikke goed°* (.) ja hoor
         °really good°       (.) yeah (hoor)
         °that's great°
   Doc:                    *nods
```

When asking how the wound is healing (line 2), the physician points at the patient's belly, indirectly orienting to the show-ability of the wound. After a single-item positive assessment by the patient's partner (line 3) and one from the patient (line 4), the patient rises to initiate a showing (Figure 4). Hence, he expands the initial positive assessments allowing the physician to independently self-assess the wound and arrive at a concurring assessment. That is, the patient's claim about the showable wound makes the showing relevant as a way of facilitating independent access (cf [24]). As the showing is emerging nonverbally, the physician produces a checking question ("yeah?" line 5), which elicits an epistemically downgraded assessment from the patient's partner ("we do think so" line 6) and an upgraded one from the patient's daughter who sits off-camera ("really pretty" line 7). The physician then receives the now perceptually available wound as newsworthy ("oh" [35]; line 8), touches the belly just over the scar with two fingers (Figure 5), and assesses the wound using the same lexical form as the patient ("looks neat"). Hence, the structure of a wound-assessment activity in the consultation room is opened with a physician question and expanded with a showing, which leads to a concurring assessment by the physician. The physician assessment is formatted with an evidential ("looks"), displaying direct access to the assessable.

Figure 4. Screenshot 1 for excerpt 2 (video time: 7:38).

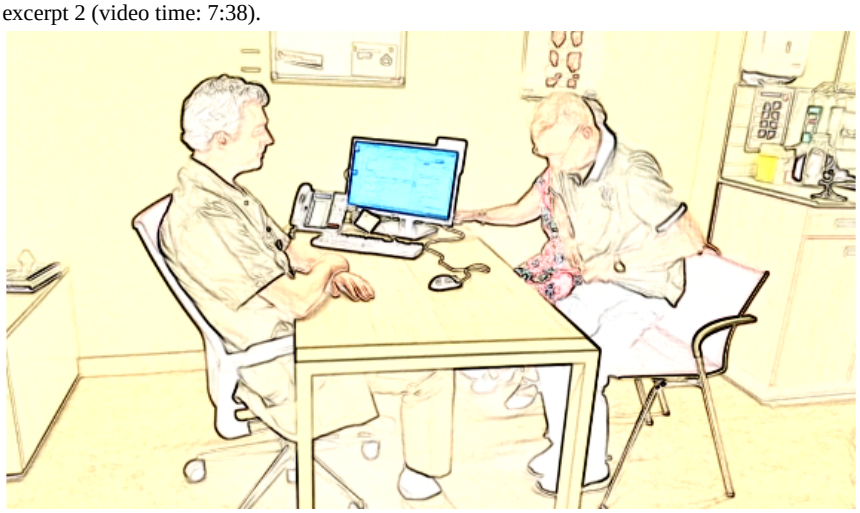
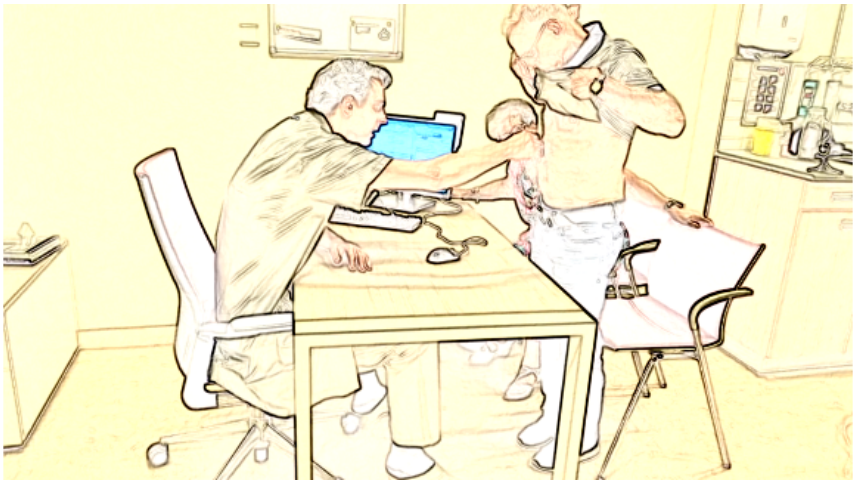


Figure 5. Screenshot 2 for excerpt 2 (video time: 7:41).



The VMC Default Method: Talk-Based Wound Assessments

The VMC default method is a talk-based assessment, which is characterized by a different structure and different epistemic marking compared to showing-based assessments. These sequences also typically begin with the physician asking whether the wound is healing well. Figure 6 is an example of a talk-based wound assessment typical for the video setting.

Figure 6. Excerpt 3. Video consultation number 16 (video time: 3:53).

1	DOC:	en U:UHM zijn- is de WOND goed genez↑en (.) de wondjes,
		and U:UHM have- has the WOUND healed well (.) the small wounds, (0.9)
2	PAT:	JA (.) honderd procent;
		YEAH (.) a hundred percent;
3		(0.8)
4	DOC:	geen uh ROODheid of ZWELling,
		no uh REDness or SWELLing,
5		(1.2)
6	PAT:	nee nee nee nee nee (.) ik kan niks opnoemen;
		no no no no no (.) I can't name anything;
7		(.)
8	DOC:	N↓EE (.) okE (.) perfect;
		NO (.) okAY (.) perfect;
9		(1.1)
10		.hh u::hm (.) NOU dat klinkt allemaal heel erg GOED;
		.hh u::hm (.) WELL that sounds all really very WELL;

The assessment activity opening question received a positive assessment from the patient, in this case with the epistemic upgrade "a hundred percent" (line 2), which formulates the confirmation as an extreme case and, thus, legitimizes it [36]. The physician does not continue with a showing request nor does the patient initiate a showing. Rather, the physician poses a subsequent question to verify the patient's answer. This question explicates two diagnostic criteria for wound assessments, namely, "redness" and "swollenness," and makes relevant a confirmation of the absence of these symptoms from the patient. The patient then responds with multiple "no's," responding to not only the immediately preceding question but to the physician's course of action, checking recovery [37]. This is elaborated with a more explicit assessment by the patient ("can't name anything"), which again legitimizes the multiple sayings of "no" (cf [36]). The physician accepts and evaluates this answer ("perfect"), and then explicitly closes the "recovery" sequence with a qualified assessment: ".hh u::hm (.) WELL that sounds all really very WELL." Note that this qualification displays the differential access [23] by the patient and physician to the wound ("sounds"), and acknowledges the patient's evaluation(s) as the epistemic basis for this closing assessment. Hence, a talk-based wound assessment in VMC is an assessment activity similar to a showing-based assessment, but it involves questioning rather than showing. Furthermore, its climax assessment reflects differential access and is, thus, epistemically weaker than in a showing-based assessment.

Even when patients produce slightly less overtly rhetoric wound assessments than "hundred percent," showings are not oriented to relevant next actions. This can be seen in Figure 7, in which the patient reports a potential minor problem with the wound ("only near my navel").

Figure 7. Excerpt 4. Video consultation number 12 (video time: 3:10).

1	DOC:	de WONDjes zijn ook goed geNEzen, **did the little WOUnds also hEAl well,**
2		(0.7)
3	PAT:	ja (.) ja de blauwe PLEkken zijn weg (.) BIJna= **yeah (.) yeah the bruises are gone (.) ALmost=**
4	DOC:	=oke= **=okay=**
5	PAT:	=en uh de wondjes zijn uh¡ (.) zijn ook GOED aan het genezen **=and uh the little wounds are uh (.) are also healing WELL**
6		dus dat zie je BIJNa niet meer sommige zie je niet meer; **so you almost don't see them anymore some you don't see anymore**
7		(0.3)
8	DOC:	[okee] **[okay]**
9	PAT:	[al]leen bij mijn NAvel; **[on]ly near my NAvel;**
10		(0.3)
11	DOC:	ja= **yeah=**
12	PAT:	=die is nog IEtsjes maar de RESt is weg; **=that one is still a BIt but the RESt is gone;**
13		(1.2)
14	DOC:	nou dat is allemaal FIJN om te HOren (.) [u:hm] **well that is all NICE to HEAr (.) [u:hm]**
15	PAT:	[ja] **[yeah]**
16	DOC:	dus uh ja dan is het toch allemaal GOED verLOpen; **so uh yeah then it did all go WELL;**

In response to the activity-opening question (line 1), the patient first confirms, produces a general description of what can be seen on the body ("bruises are gone (.) almost"), and then gives an explicit verbal confirmation that the wounds are healing well. This is elaborated with a formulation of the visual observation of the wounds as the epistemic basis of this claim ("you almost don't see them anymore some you can't see anymore" line 6). Note that the patient uses the impersonal "you," designing this claim as objective rather than as epistemically marked as her own observation.

The patient then expands the positive assessment with a minor problem ("only near my navel" line 9, "that one is still a bit" line 12) although this is contrasted with an overall positive assessment ("but the rest is gone" line 12), which proposes a closing of the assessment. The physician responds with the qualified assessment "well that is all NICE to Hear," not orienting to the minor problem report but treating the patient's wound assessment as relatively unspecific ("all") and as news that he had no direct independent access to. The presented conclusion that follows ("so uh yeah then it did all go WELL") is built on this general, positive news receipt and covers the whole surgery process, thus, moving out of the activity of wound assessment.

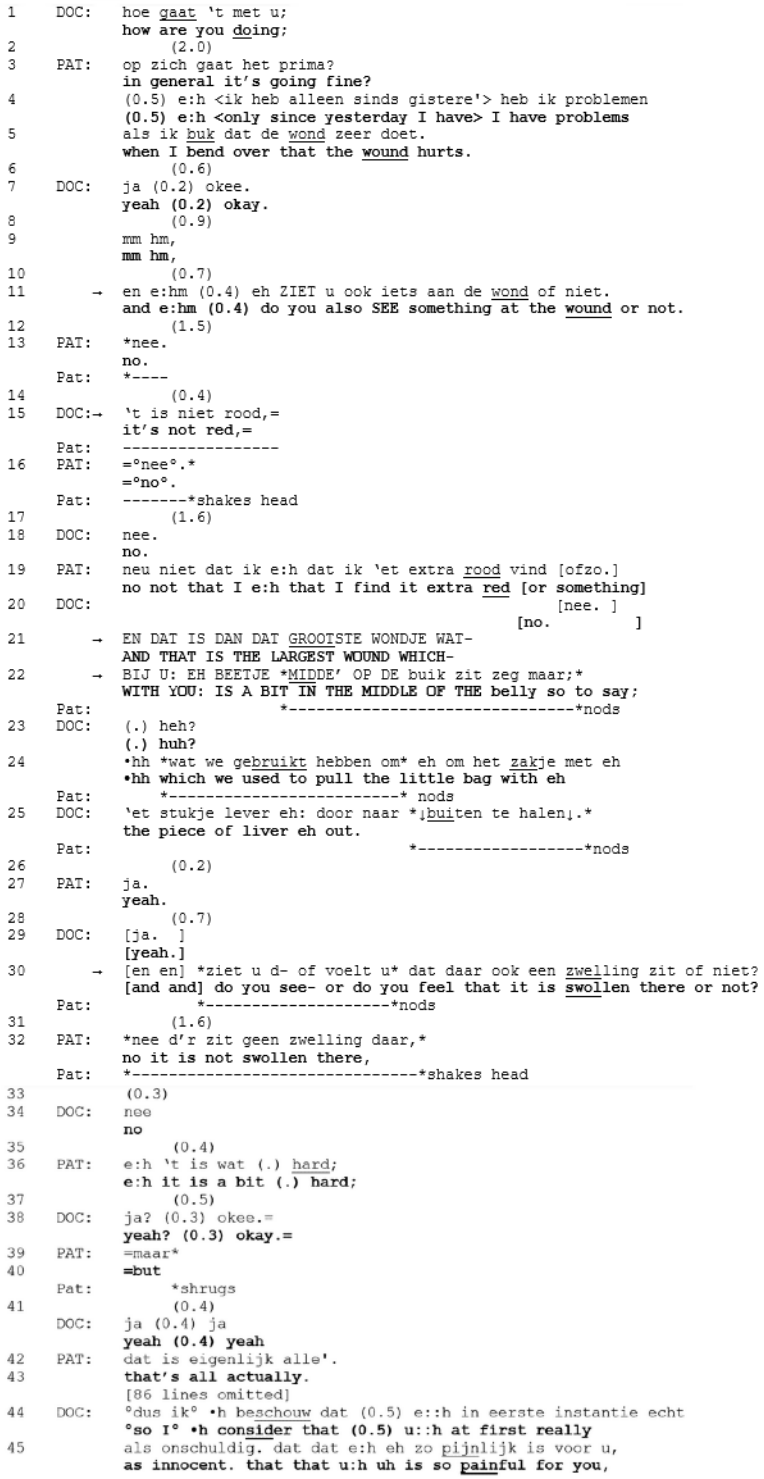
In summary, talk-based wound assessment sequences include the specification of diagnostic criteria ("redness," "swollenness"), perceptual basis ("some you don't see anymore"), or reference to a location on the body ("near my navel"). These may be elicited by the physician or volunteered by the patient. Generally, in such VMC talk-based assessments, physicians arrive at qualified wound assessments, marking them as epistemically grounded in the patient's evaluation rather than in their own observation or examination.

Patient-Initiated Wound Assessments in VMC

In the examples so far, the assessment sequences were initiated by the physician enquiring about the wound. However, wound assessments may also be initiated by patients rather than physicians with a report or question addressing some sort of trouble with regard to the wound. Although wound (or location) showings do occur (n=3/22 of video consultations), even patient-initiated sequences, which make a wound assessment relevant, may unfold as *talk-based* assessments in video consultations. In such cases, the interaction tends to be stretched over several sequences. The possibility of showing the wound is disregarded, despite the fact that visual access is available through the video connection. Figure 8 shows an example of a lengthy talk-based wound assessment in a video setting. In response to the question about how the patient is doing (line 1), she reports pain related to the wound as a first concern (cf [13]), which opens up the wound assessment activity.

To begin with, the patient refers to the viewable wound in her presentation of the problem (lines 4-5), which creates an opportunity for the physician to request a showing (a so-called "touched-off" showing [38]). However, rather than requesting that the patient show the wound, he accepts this initial problem account ("yeah (0.2) okay" line 7), which arguably projects history taking as a next activity [39]. Nevertheless, the physician does *not* take a turn, remains silent for 0.9 seconds, and then produces a continuer ("mhmm" line 9), creating a context for the patient to elaborate on the complaint. In the silence that follows (0.7 seconds), the patient does not continue, and the physician initiates the talk-based assessment activity. Hence, the initial problem report by the patient seems to create an interactional limbo in the structure of the assessment activity in which an opportunity to request a showing has passed.

Figure 8. Excerpt 5. Video consultation 1 (video time: 1:54).



The physician, asking whether the patient also sees something at the wound, orients to the viewability and, thus, potential show-ability of the wound. The patient denies something can be seen, and the physician expands by making explicit what could be seen, namely, redness (line 15). This further question also receives a "no," but is then elaborated on with an account that implicitly proposes a scale of redness indicating the wound is not "extra red" (line 19). "Not extra red or so" implies the wound is (a bit) red, which may be a flag for trouble to the physician. Saliently, the patient produces this assessment without direct visual access (ie, she is not simultaneously inspecting the wound). Moreover, it is marked with an epistemic downgrade ("I find" line 19), thus, designing this observation as not only rooted in her earlier observation but also as subjective (or "subject-side" [40]). A showing could have resolved these issues, but this is not what happens. Rather, the physician checks whether the complaint is about the largest wound (line 21) and indicates roughly where this is located on the patient's belly (line 22). In this way, again, a viewable (location on the belly) is talked about without being shown. After the participants have reached an agreement about which wound is being discussed (line 27 and 29), the physician launches another diagnostic

question: "[and and] do you see- or do you feel that it is swollen there or not?" The seeing as a source of observation is redirected to feeling regarding the diagnostic criterion of "swollenness" (line 30). Hence, the patient is not invited to look "through the doctor's eyes," but to touch "on behalf of the physician." This implies showing is now less relevant, as a showing might be perceptually inadequate to assess swollenness.

From this point onward, the patient reports tactile observations including that it is "not swollen there" (line 32), followed by a further potentially troublesome description "it's a bit (.) hard" (line 36), which introduces yet another category. Nevertheless, the sequence is collaboratively closed with an orientation to the problem as minimal ("that's all actually" line 43) and, thus, not in need of further discussion. The physician then starts a new but related sequence on the patient's activities during the past weeks (data not shown), which eventually leads to his wound and pain assessment as "innocent" (line 45) with multiple disfluencies and hedges ("uh's," "at first really"), and an epistemic downgrade ("I consider that"). Hence, a talk-based assessment in cases of potential trouble may reside in talk to avoid a showing request. It includes the explication of multiple diagnostic criteria and may involve reference to various sensorial observations by the patient, and it eventually leads to a qualified wound assessment.

Discussion

Principal Results

Our primary finding is that video consultations differ from copresent consultations with regard to wound assessment. Talk-based wound assessment is the dominant trajectory in video consultations, while showing-based wound assessment is the dominant method in copresent consultations. Both trajectories are generally initiated with an informing question by the physician, but the subsequent steps differ. The activity continues with either a showing or examination of the wound, or with one or more questions enquiring the absence of specific diagnostic criteria (eg, redness, swollenness). Showing-based assessments work toward evidentially grounded general assessments ("neat," "good," "uneventful"), while talk-based assessments arrive at qualified assessments, which display a lack of direct access to the assessable ("sounds," "I consider that"). Hence, wound assessments in video consultations are grounded in patient assessments, which implies a shift in clinical practice from primacy of the doctor's gaze to the patient's evaluation of how the wound(s) are healing. Even in cases of potential wound trouble in video consultations, physicians may rely on talk and avoid requesting a showing of the wound despite its apparent relevance. Such talk-based assessment sequences can be stretched substantially, with physicians bringing up multiple questions to enquire about symptoms and observations from the patient, both visual and tactile. Hence, despite the possibility of visual access and the interactional relevance, the participants displayed an orientation to avoid a showing in video consultations.

Comparison With Prior Work

We may speculate about the reasons for the avoidance of showing closed surgical wounds in video consultations. A general reason could be that asking a patient to undress or show part of the nude belly or torso while being in the private sphere (usually the living room) with others potentially present and showing part of the nude body on camera are delicate things to do. In contrast, the hospital's consultation room is marked with a clinical setup and assets (eg, physician wearing white coat, curtain, examination table, medical instruments), creating a context where showing the body and physical examinations may become relevant or may be expected by patients or physicians. Possibly, as participants' experience with video interactions evolve, showing practices may occur more naturally. The avoidance of showings and, thus, direct visual access by the physician in video consultations implies that the "ecological advantage" [16] of physical examination may not or does not naturally apply to the video setting.

This means that physicians have less authority in diagnosing the wound and that patients are instead more agentive and epistemically amplified compared to face-to-face consultations. Similarly, Seuren et al [10] suggested that physical examination in video consultations may enhance patient autonomy, as patients become more active participants in the examination, having to handle instruments (eg, to measure oxygen in blood) or modify the camera. Nevertheless, physicians' qualified assessments indicate a degree of uncertainty as a result of the restrictions of the medical armentarium inherent to the medium (eg, the impossibility of palpation) and reliance on patient reports and observations (cf [12]). This might explain why prior studies on physicians' perspectives on applicability of video consultations revealed an anticipated need for physical examination as the main reason for not opting for video consultations [1,6]. The question is, however, under which conditions is it necessary to conduct a physical examination. In the majority of cases, talk-based assessment was sufficient to assess wound-healing, which implies the early postoperative phase is a context in which video consultations appear effective.

Limitations

A limitation of our study is that we cannot exclude that the patients who chose a hospital consultation were more insecure about their recovery, including the wound(s), than those who opted for a video consultation. In that case, our findings could not only be explained by the medium of communication. However, in examining the data, we found multiple cases of patients in the hospital setting who did not present any insecurity with regard to their recovery, and we also found cases of potential insecurity (ie, patient reporting pain) in the video data. Another limitation is that the observed phenomenon may be related to the specific goal of the consultation. In our data, the reason for the consultation was the news delivery of the pathology results, and an examination of the wound was not explicitly announced. However, in the face-to-face consultations, showings and the physicians' invitations to "have a look" were utterly unproblematic. Nevertheless, it is possible that in video consultations where the goal of the interaction is more closely linked to examination, medically relevant showings are more common and are also volunteered by patients (cf [10,12]). Patients may even close the curtains or do the video consultation from their bedroom. Hence, medical assessment practices are

likely to further evolve with participants' growing familiarity with video-mediated interactions.

Conclusions

Overall, it has become clear that video-mediated and copresent medical interactions differ with regard to assessments of medical assessables such as wounds. It was particularly the comparative perspective that yielded new insights, providing evidence of normative orientations with regard to showing that intersect the medical dimension of the talk and the medium of communication. This underscores the relevance of the communication channel for the organization of institutional talk-in-interaction [41] and shows that juxtaposing equivalent interactions through different media is worthwhile, particularly when the choice for the one or the other medium is an "emic" choice for both the patient and—in a different way—the physician. New communicative affordances such as patients sharing images of their body parts with their phones are likely to further affect medical interactions.

A practical implication of our study is that physicians may have to do "extra work" in video consultations to facilitate showing-based assessments. Furthermore, they should consider under which circumstances (eg, closed wound inspection) a hospital visit is more suitable than a video consultation. Another practical implication is that talk-based assessment seems to reduce the physician's medical authority, as it ascribes more authority to the patient. This reliance on patient observation and judgement is in line with increased self-management as a form of patient empowerment in video consultations [10,42], which is generally regarded as beneficial. It nevertheless seems important that practitioners are aware of potentially shifting authority.

Acknowledgments

We would like to thank Christian Licoppe for a data session at Télécom Paristech and the participants of data sessions at Copenhagen University and at King's College London for their valuable input to the analysis of our data. We are also grateful to Willem van de Heuvel and the computer and information sciences MA students for their initial transcriptions of the videos, Anita Pomerantz for her feedback on an earlier version of the manuscript, and to two anonymous reviewers for their valuable comments.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Transcription conventions.
[PDF File (Adobe PDF File), 104 KB - jmir_v22i5e17791_app1.pdf]

References

1. Donaghy E, Atherton H, Hammersley V, McNeilly H, Bikker A, Robbins L, et al. Acceptability, benefits, and challenges of video consulting: a qualitative study in primary care. Br J Gen Pract 2019 Jun 03;69(686):e586-e594. [doi: 10.3399/bjgp19x704141]
2. Di Cerbo A, Morales-Medina JC, Palmieri B, Iannitti T. Narrative review of telemedicine consultation in medical practice. Patient Prefer Adherence 2015;9:65-75. [doi: 10.2147/PPA.S61617] [Medline: 25609928]
3. Sabesan S, Kelly J, Evans R, Larkins S. A tele-oncology model replacing face-to-face specialist cancer care: perspectives of patients in North Queensland. J Telemed Telecare 2014 Jun;20(4):207-211. [doi: 10.1177/1357633X14529237] [Medline: 24643950]
4. Viers BR, Lightner DJ, Rivera ME, Tollefson MK, Boorjian SA, Karnes RJ, et al. Efficiency, satisfaction, and costs for remote video visits following radical prostatectomy: a randomized controlled trial. Eur Urol 2015 Oct;68(4):729-735. [doi: 10.1016/j.eururo.2015.04.002] [Medline: 25900782]
5. Thiyagarajan A, Grant C, Griffiths F, Atherton H. Exploring patients' and clinicians' experiences of video consultations in primary care: a systematic scoping review. BJGP Open 2020 Apr 03. [doi: 10.3399/bjgpopen20X101020] [Medline: 32184212]
6. Greenhalgh T, Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, et al. Real-World Implementation of Video Outpatient Consultations at Macro, Meso, and Micro Levels: Mixed-Method Study. J Med Internet Res 2018 Apr 17;20(4):e150 [FREE Full text] [doi: 10.2196/jmir.9897] [Medline: 29625956]
7. Barsom EZ, Jansen M, Tanis PJ, van de Ven AWH, Blussé van Oud-Alblas M, Buskens CJ, et al. Video consultation during follow up care: effect on quality of care and patient- and provider attitude in patients with colorectal cancer. Surg Endosc 2020 Mar 20. [doi: 10.1007/s00464-020-07499-3] [Medline: 32198552]
8. Atherton H, Brant H, Ziebland S, Bikker A, Campbell J, Gibson A, et al. The potential of alternatives to face-to-face consultation in general practice, and the impact on different patient groups: a mixed-methods case study. NIHR Journals Library 2018 Jun. [doi: 10.3310/hsdr06200] [Medline: 29889485]

9. Ignatowicz A, Atherton H, Bernstein CJ, Bryce C, Court R, Sturt J, et al. Internet videoconferencing for patient-clinician consultations in long-term conditions: a review of reviews and applications in line with guidelines and recommendations. Digit Health 2019;5:2055207619845831 [FREE Full text] [doi: 10.1177/2055207619845831] [Medline: 31069105]
10. Seuren LM, Wherton J, Greenhalgh T, Cameron D, A'Court C, Shaw SE. Physical Examinations via Video for Patients With Heart Failure: Qualitative Study Using Conversation Analysis. J Med Internet Res 2020 Feb 20;22(2):e16694 [FREE Full text] [doi: 10.2196/16694] [Medline: 32130133]
11. Pappas Y, Seale C. The opening phase of telemedicine consultations: an analysis of interaction. Soc Sci Med 2009 Apr;68(7):1229-1237. [doi: 10.1016/j.socscimed.2009.01.011] [Medline: 19201514]
12. Pappas Y, Seale C. The physical examination in telecardiology and televascular consultations: a study using conversation analysis. Patient Educ Couns 2010 Oct;81(1):113-118. [doi: 10.1016/j.pec.2010.01.005] [Medline: 20144523]
13. Stommel W, Van Goor H. Other-attentiveness in video consultation openings: a conversation analysis of video-mediated versus face-to-face consultations. Journal of Computer-Mediated Communication 2019:15. [doi: 10.1093/jcmc/zmz015]
14. Hutchby I. Conversation and technology: From the Telephone to the Internet. Cambridge: Polity Press; 2001.
15. Heath C, Luff P. Disembodied conduct: interactional asymmetries in video-mediated communication. In: Button G, editor. Technology in Working Order: Studies of Work, Interaction, and Technology. London: Routledge; 1993:35-54.
16. Heritage J. Online commentary in primary care and emergency room settings. Acute Med Surg 2017 Jan;4(1):12-18. [doi: 10.1002/ams2.229] [Medline: 29123831]
17. Heritage J, Stivers T. Online commentary in acute medical visits: a method of shaping patient expectations. Social Science & Medicine 1999 Dec;49(11):1501-1517. [doi: 10.1016/s0277-9536(99)00219-1]
18. Mangione-Smith R, Stivers T, Elliott M, McDonald L, Heritage J. Online commentary during the physical examination: a communication tool for avoiding inappropriate antibiotic prescribing? Social Science & Medicine 2003 Jan;56(2):313-320. [doi: 10.1016/s0277-9536(02)00029-1]
19. Perakyla A. Authority and accountability: the delivery of diagnosis in primary health care. Social Psychology Quarterly 1998 Dec;61(4):301. [doi: 10.2307/2787032]
20. Goodwin C. Professional vision. American Anthropologist 1994 Sep;96(3):606-633. [doi: 10.1525/aa.1994.96.3.02a00100]
21. Lindström A, Mondada L. Assessments in social interaction: introduction to the special issue. Research on Language & Social Interaction 2009 Nov 25;42(4):299-308. [doi: 10.1080/08351810903296457]
22. Pomerantz A. Agreeing and disagreeing with assessments: some features of preferred/dispreferred turn shapes. In: Atikson JM, Heritage J, editors. Structures of Social Action: Studies in Conversation Analysis. Cambridge: Cambridge University Press; 1984:55-101.
23. Goodwin C, Goodwin MH. Assessments and the construction of context. In: Rethinking Context: Language as an Interactive Phenomenon. Cambridge: Cambridge University Press; 1992:147-190.
24. Licoppe C. Showing objects in Skype video-mediated conversations: from showing gestures to showing sequences. Journal of Pragmatics 2017 Mar;110:63-82. [doi: 10.1016/j.pragma.2017.01.007]
25. Fasulo A, Monzoni C. Assessing mutable objects: a multimodal analysis. Research on Language & Social Interaction 2009 Nov 25;42(4):362-376. [doi: 10.1080/08351810903296481]
26. Oshima S, Streeck J. Coordinating talk and practical action: the case of hair salon service assessments. PS 2015 Dec 7;6(4):538-564. [doi: 10.1075/ps.6.4.04osh]
27. Raclaw J, Robles JS, DiDomenico SM. Providing epistemic support for assessments through mobile-supported sharing activities. Res Lang Soc Interact 2016;49(4):362-379 [FREE Full text] [doi: 10.1080/08351813.2016.1199089] [Medline: 28936031]
28. Olbertz-Siitonen M. Transmission delay in technology-mediated interaction at work. PsychNology Journal 2015:203.
29. Licoppe C, Morel J. Video-in-interaction: “talking heads” and the multimodal organization of mobile and Skype video calls. Research on Language & Social Interaction 2012 Oct;45(4):399-429. [doi: 10.1080/08351813.2012.724996]
30. Jefferson G. Glossary of transcript symbols with an introduction. In: Lerner G, editor. Conversation Analysis: Studies From the First Generation. Amsterdam and Philadelphia: John Benjamins; 2004:14-31.
31. Mondada L. Multiple temporalities of language and body in interaction: challenges for transcribing multimodality. Research on Language and Social Interaction 2018 Mar 09;51(1):85-106. [doi: 10.1080/08351813.2018.1413878]
32. Sidnell J, Stivers T. The Handbook of Conversation Analysis. London: Blackwell Publishing; 2013.
33. Stivers T, Sidnell J. Introduction: multimodal interaction. Semiotica 2005:1. [doi: 10.1515/semi.2005.2005.156.1]
34. Curl T, Drew P. Contingencey and action: a comparison of two forms of requesting. Reseach on Language and Social Interaction 2008:1.
35. Heritage J. A change-of-state tokenaspects of its sequential placement. In: Atkinson M, Heritage J, editors. Structures of Social Action. Cambridge: Cambridge University Press; 1984:299-245.
36. Pomerantz A. Extreme case formulations: a way of legitimizing claims. Hum Stud 1986;9(2-3):219-229. [doi: 10.1007/BF00148128]
37. Stivers A. "No no no" and other types of multiple sayings in social interaction. Human Comm Res 2004 Apr;30(2):260-293. [doi: 10.1111/j.1468-2958.2004.tb00733.x]

38. Licoppe C, Tuncer S. The initiation of showing sequences in video-mediated communication. Discourse and Conversation Analysis 2020.
39. Robinson JD. An interactional structure of medical activities during acute visits and its implications for patients' participation. Health Communication 2003 Jan;15(1):27-59. [doi: 10.1207/s15327027hc1501_2]
40. Wiggins S, Potter J. Attitudes and evaluative practices: category vs. item and subjective vs. objective constructions in everyday food assessments. Br J Soc Psychol 2003 Dec;42(Pt 4):513-531. [doi: 10.1348/014466603322595257] [Medline: 14715115]
41. Stommel W, te Molder H. Counseling online and over the phone: when preclosing questions fail as a closing device. Research on Language and Social Interaction 2015 Aug 11;48(3):281-300. [doi: 10.1080/08351813.2015.1058605]
42. Hinman RS, Nelligan RK, Bennell KL, Delany C. "Sounds a Bit Crazy, But It Was Almost More Personal:" a qualitative study of patient and clinician experiences of physical therapist-prescribed exercise for knee osteoarthritis via Skype. Arthritis Care Res (Hoboken) 2017 Dec;69(12):1834-1844. [doi: 10.1002/acr.23218] [Medline: 28217864]

Abbreviations

VMC: video-mediated communication

Edited by G Eysenbach; submitted 13.01.20; peer-reviewed by L Seuren, T Halkowski; comments to author 10.03.20; revised version received 03.04.20; accepted 08.04.20; published 05.05.20.

Please cite as:
Stommel WJP, van Goor H, Stommel MWJ
The Impact of Video-Mediated Communication on Closed Wound Assessments in Postoperative Consultations: Conversation Analytical Study
J Med Internet Res 2020;22(5):e17791
URL: https://www.jmir.org/2020/5/e17791
doi:10.2196/17791
PMID:32310816

Original Paper

Video Consultations Between Patients and Clinicians in Diabetes, Cancer, and Heart Failure Services: Linguistic Ethnographic Study of Video-Mediated Interaction

Sara E Shaw[1], PhD; Lucas Martinus Seuren[1], PhD; Joseph Wherton[1], PhD; Deborah Cameron[2], MLitt; Christine A'Court[1], MD; Shanti Vijayaraghavan[3], MD; Joanne Morris[3], PhD; Satyajit Bhattacharya[3], FRCS; Trisha Greenhalgh[1], MD

[1]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

[2]Faculty of Linguistics, University of Oxford, Oxford, United Kingdom

[3]Barts Health NHS Trust, London, United Kingdom

Corresponding Author:
Sara E Shaw, PhD
Nuffield Department of Primary Care Health Sciences
University of Oxford
Radcliffe Observatory Quarter
Woodstock Road
Oxford, OX2 6GG
United Kingdom
Phone: 44 1865617830
Email: sara.shaw@phc.ox.ac.uk

Abstract

Background: Video-mediated clinical consultations offer potential benefits over conventional face-to-face in terms of access, convenience, and sometimes cost. The improved technical quality and dependability of video-mediated consultations has opened up the possibility for more widespread use. However, questions remain regarding clinical quality and safety. Video-mediated consultations are sometimes criticized for being not as good as face-to-face, but there has been little previous in-depth research on their interactional dynamics, and no agreement on what a good video consultation looks like.

Objective: Using conversation analysis, this study aimed to identify and analyze the communication strategies through which video-mediated consultations are accomplished and to produce recommendations for patients and clinicians to improve the communicative quality of such consultations.

Methods: We conducted an in-depth analysis of the clinician-patient interaction in a sample of video-mediated consultations and a comparison sample of face-to-face consultations drawn from 4 clinical settings across 2 trusts (1 community and 1 acute care) in the UK National Health Service. The video dataset consisted of 37 recordings of video-mediated consultations (with diabetes, antenatal diabetes, cancer, and heart failure patients), 28 matched audio recordings of face-to-face consultations, and fieldnotes from before and after each consultation. We also conducted 37 interviews with staff and 26 interviews with patients. Using linguistic ethnography (combining analysis of communication with an appreciation of the context in which it takes place), we examined in detail how video interaction was mediated by 2 software platforms (Skype and FaceTime).

Results: Patients had been selected by their clinician as *appropriate* for video-mediated consultation. Most consultations in our sample were technically and clinically unproblematic. However, we identified 3 interactional challenges: (1) opening the video consultation, (2) dealing with disruption to conversational flow (eg, technical issues with audio and/or video), and (3) conducting an examination. Operational and technological issues were the exception rather than the norm. In all but 1 case, both clinicians and patients (deliberately or intuitively) used established communication strategies to successfully negotiate these challenges. Remote physical examinations required the patient (and, in some cases, a relative) to simultaneously follow instructions and manipulate technology (eg, camera) to make it possible for the clinician to see and hear adequately.

Conclusions: A remote video link alters how patients and clinicians interact and may adversely affect the flow of conversation. However, our data suggest that when such problems occur, clinicians and patients can work collaboratively to find ways to overcome them. There is potential for a limited physical examination to be undertaken remotely with some patients and in some

conditions, but this appears to need complex interactional work by the patient and/or their relatives. We offer preliminary guidance for patients and clinicians on what is and is not feasible when consulting via a video link.

International Registered Report Identifier (IRRID): RR2-10.2196/10913

(J Med Internet Res 2020;22(5):e18378) doi:10.2196/18378

KEYWORDS

delivery of health care; physical examination; remote consultation; telemedicine; health communication; language; nonverbal communication; mobile phone

Introduction

Background

There is a significant push from decision makers across the world to make better use of digital technologies, including video consultations [1-3]. Virtual media such as Skype (Microsoft Corporation), FaceTime (Apple Inc), and Attend Anywhere (Attend Anywhere) are increasingly being used for web-based communication between patients and clinicians. In outpatient departments, drivers include addressing high nonattendance rates and rising costs. Patients are often keen to use video consultations given the potential for reduced time and travel, especially for tertiary care [4-6]. The coronavirus disease 2019 (COVID-19) pandemic has provided significant additional pressure to deliver video-mediated services, rapidly and at scale, to reduce face-to-face social contact and help to reduce the spread of the disease.

Published research on video outpatient consultations has been summarized in several recent reviews [7-10]. These show great potential for video consultations in terms of acceptability, safety, and effectiveness in patients considered clinically eligible across a range of conditions, such as diabetes [11-15], ophthalmology [16], cancer [17,18], chronic kidney disease [19], spinal cord injury [6,20,21], chronic obstructive pulmonary disease [22,23], mental health conditions [24,25], Down syndrome [26], cerebral palsy [27], chronic pain [18,28,29], therapies (eg, speech and language therapy [30-32]), support after premature birth [33], support of patients in care homes [34], and plastic surgery [35].

Much of this literature uses either experimental methods, classifying service models primarily by the technology used and secondarily by the task or clinical work supported by that technology, or questionnaires and semistructured interviews, typically with small samples, to assess feasibility and acceptability. There is limited evidence on costs [36]. Patients have expressed some concerns about privacy and security and ease of use of the technology [37,38]. Some studies have reported less enthusiasm for video consultations among older people [39,40]. Qualitative studies have reported concerns over technical issues such as pixilation or audio quality and quality of care, for instance, because of the lack of physical examination [9,38,41,42].

There has been limited research focused specifically on the dynamics of video-mediated interaction in health care consultations and the ways in which communication might be altered by the use of video. To date, only 8 studies have examined such interaction in video consultations, with only one analyzing video recordings at both *ends* of the consultation. Ekberg et al [30,43] looked at the use of virtual media to support cognitive behavioral therapy and speech and language therapy, finding that in both cases, practitioners were able to adapt consultations and communication accordingly. In a series of studies on video-mediated vascular and cardiology consultations, Pappas et al [44-46] explored video consultations in which the patient, supported by a general practitioner (GP) or nurse, consulted remotely with a specialist from their GP practice. They found that clinical assessment and decision making worked well and that the video medium supported collaborative discussion and shared care. However, professionals needed to adapt their interaction to accommodate the video medium (eg, switching between consultative and interprofessional talk, while manipulating camera angle) so as to ensure ongoing patient involvement [44-46]. A previous study by our team compared video consultations of diabetic and postoperative cancer patients with matched face-to-face consultations and provided an overview of what is gained and what is lost when clinicians and patients interact remotely for instance, finding that the overall length of a video consultation is shorter and there are more breaks in the conversation compared with an equivalent face-to-face encounter [9,47]. Stommel et al [48] studied openings of postsurgical video consultations, revealing small but significant interactional differences, with more prosocial talk at the start of video consultations, and the role of companions in a video consultation, revealing that they are mostly off-screen, invisible to the clinician, and rarely involved in the interaction [49].

There is a growing body of evidence outside of health care that highlights the ways in which video calling might alter interaction [50]. Early studies on video-mediated interaction highlighted how technology could be detrimental to conversational flow [51-53]. In face-to-face interaction, people can see and hear actions as they are produced, but in video-mediated interaction, they see and hear them milliseconds later, a phenomenon known as latency. Such delays seem small but are meaningful in conversation as they interfere with the system of rules that participants use to determine who gets to talk at which point in the conversation (turn-taking [54]) and address problems such as misunderstandings (repair [55]). By preventing participants from adequately applying these rules, video technology can affect the quality of interaction and user experience. Other studies raise questions about if technical failures, such as hearing no sound or new types of interruption, for instance, a family member entering the room, impact the interaction [56-58]. Such evidence has yet to be considered in relation to video consultations in health care.

In summary, although the evidence on video consultations indicates that they are feasible, safe, and effective in health care, the evidence on interaction in such consultations is limited. Only a handful of studies have considered the impact of the video medium on the microdynamics of interaction in medical consultations [36]. The result is that little is currently known about how different communication strategies, types of communication (speech, bodily conduct, gaze, and posture), the material properties of technology, and/or the quality of web-based connection shape interaction in video consultations in health care.

We studied interaction in video consultations in four clinics in two National Health Service (NHS) trusts (one acute and one community). The Qualitative Analysis of Remote Consultations (QuARC) study built on several years' work by our team looking at the national-level context for video consultations, the organizational context supporting adoption and implementation, and microlevel interactions between patients and clinicians [9,47,59]. The QuARC study combined data from 2 previous studies—the Virtual Online Consultations-Advantages and Limitations (VOCAL) study, a National Institute for Health Research (NIHR)–funded study of diabetes and cancer clinics in East London using Skype, and the Oxford Telehealth Qualitative Study (OTQS), a Wellcome Trust-funded study of specialist community heart failure teams in Oxford using Skype or FaceTime—and conducted detailed interactional microanalysis across video consultation data.

Aims, Objectives, and Research Questions

As set out in the QuARC study protocol [10], the aims of the study were to identify and analyze the communication strategies through which remote consultations are accomplished and to produce guidance for patients and clinicians to improve the communicative quality of video consultations. Specifically, our objectives were (1) to conduct a secondary analysis of a multimodal dataset of 37 remote consultations with diabetes, antenatal diabetes, cancer, and heart failure patients and their clinicians using a combination of ethnographic and microanalytic approaches to investigate how interaction is affected by mediation via Skype or similar apps and (2) develop guidance for patients and clinicians in conducting remote consultations.

Our research questions were as follows:

1. What are the (often implicit or unspoken) communication strategies through which technology-mediated consultations for diabetes, cancer, and heart disease are successfully accomplished?
2. How do patients and clinicians negotiate misunderstandings in technology-mediated consultations, and what strategies are more effective?
3. What can we learn from a detailed linguistic analysis of real-life remote consultations to guide other clinicians and patients interested in or actively using Skype and other video platforms for medical consultations?

Methods

We have previously published a detailed study protocol [10]. This section provides a summary along with an update on the analytic approach.

Study Design and Approach

A mixed methods design was used to study video consultations in 4 contrasting clinical settings (diabetes, antenatal diabetes, cancer, and heart failure patients) across 2 NHS trusts (Oxford and London). The study was informed by linguistic ethnography, which uses both linguistic and ethnographic approaches to understand how social and communicative processes operate in a range of settings and contexts, combining video and audio recordings of consultations, interviews, and observations from 2 previous studies (VOCAL and OTQS, detailed above) to examine in detail how patient-clinician interaction is shaped by use of Skype and FaceTime.

Project Management and Governance

This study was delivered by a core working group (SS, LS, DC, JW, and TG) and supported by an independent steering group with a lay chairperson and cross-sector stakeholder representation, including patients, NHS stakeholders, and national-level decision makers.

Both studies from which data were drawn received ethical approval for a detailed analysis of video recordings of video consultations. VOCAL was approved by the National Research Ethics Service Committee London–City Road and Hampstead in December 2014 (14/LO/1883) and OQTS by the South Central–Berkshire Research Ethics Committee in September 2015 (15/SC/0553). All participating staff and patients in both studies gave their informed consent to be audio and video recorded during consultations and for the data to be used for research purposes.

Setting

The study took place at 2 NHS trusts in London and Oxford, each of which explored the use of video consultation services. Table 1 provides an overview of the clinical settings in which recordings of consultations took place. The diabetes service in London had an established video consultation service, whereas the other 3 were piloting or setting up video consultations at the time of the respective studies. Video consultation services in heart failure and antenatal diabetes ceased at the end of the respective studies. Further details on each of the settings can be found in publications from VOCAL [9,47] and OTQS (C Papoutsi et al, unpublished data, 2020).

Table 1. Summary of the four clinical settings in which video consultations took place.

Clinical setting	Population	Clinical provision	Staff	Video consultation service
Diabetes services (London)	Adult/young; adult patients (18-80+ years), with high prevalence of type 2 diabetes in ages 16 to 25 years, plus significant risk factors (eg, poverty, diet, or ethnicity)	Integrated community diabetes service, with consultants providing 6-monthly reviews and ongoing support from diabetes nurse specialists	Lead diabetologist, 5 consultant diabetologists, and 6 specialist nurses	Established in 2010 because of typically low engagement with traditional service models, poor health outcomes, increasing use of unplanned care via A&E[a]; delivered largely by lead diabetologist (who offered virtual consultations to all adult/young adult patients as an alternative to follow-ups), with other staff slowly coming on board; using Skype (consumer version) on desktop at the time of the study
Antenatal diabetes services (London)	Expectant mothers (around 350 per year) with gestational diabetes	Outpatient consultations (including preappointment tests and checks) combined with optional weekly telephone clinic (for those needing close monitoring); key medical information (eg, blood sugar readings) stored in patient-held maternity folder	3 diabetes consultants, 3 obstetricians, 2 nurses, and 1 midwife	Piloted as part of the VOCAL[b] study, with video consultations led by 1 consultant and using Skype (consumer version) on a clinic desktop
Hepatobiliary and pancreatic cancer surgery services (London)	Patients with pancreatic/liver cancer who had major surgery and a prolonged postoperative phase; diverse demographic, living up to 200 miles from clinic	Tertiary service, with clinic run once per week, 2 to 3 patients were typically seen for postoperative cancer follow-up	1 consultant surgeon, 2 specialist registrars, 1 clinical nurse specialist, and nurse assistants	At the start of the VOCAL study, the clinic had begun to introduce virtual consultations to spare selected patients unnecessary travel, run in a shared hospital space alongside other clinical services, and using Skype (consumer version) on a clinic desktop
Heart failure service (Oxford)	Heart failure patients (typically 65+ years) with reduced ejection fraction, many unable to get to clinic (owing to frailty or severe symptoms)	Community outreach service delivered by heart failure specialist nurses working with the hospital-based heart failure service, local general practitioners, other community services, social services, and ambulatory assessment units	5 specialist heart failure nurses	Piloted at the time of the OTQS[c] study to evaluate if video consultations could help deploy limited resources safely, efficiently, and effectively without loss of patient or staff satisfaction. Heart failure specialist nurses were equipped by their employing trust with iPads with SIM cards to enable real-time access to patients' records, enabling the use of Skype or FaceTime

[a]A&E: accident and emergency.

[b]VOCAL: Virtual Online Consultations-Advantages and Limitations.

[c]OTQS: Oxford Telehealth Qualitative Study.

Sampling and Data Collection

We collected data over 30 months (VOCAL, 2015-2017; OTQS 2016-2017). The data sources are summarized in Table 2, with further details below.

Our dataset consisted of 37 video recordings of consultations (Table 2) from the VOCAL and OTQS studies. The goal of sampling in both studies was to capture the breadth of experience of video consultations, seeking maximum variety in clinical, social, ethnic, and personal circumstances (Table 3) and in health and information technology literacy. It was a precondition of ethical approval that clinicians could exercise judgment about which patients to invite to the study.

Exclusion criteria were no 3G internet access at home, lack of familiarity with the technology, clinical inappropriateness (eg, in London, the need for direct physical examination), inability to give informed consent, and comorbidity preventing participation (eg, severe visual impairment). In the diabetes clinic, which included a high proportion of limited English speakers, bilingual health advocates trained in the use of remote consulting were available, so limited English was not an exclusion criterion there. In the antenatal diabetes and cancer surgery clinics, those who were comfortable with a family member interpreter were included, but a remote interpreting service was not available. In the heart failure service, all participating patients were native English speakers.

To enable comparison, we collected 28 audio recordings of matched face-to-face consultations (Table 2). Participants were selected on the basis that the type of consultation was similar to those conducted remotely (eg, routine follow-up cancer appointments) and that the consultant would consider such cases suitable for a video consultation.

Table 2. Overview of data and analysis of the Qualitative Analysis of Remote Consultations study.

Type of data	Data description	First order interpretation	Higher order interpretation
Consultation data	• Video recordings and screen capture (at patient end and clinician end) of 37 virtual consultations (12 diabetes, 6 antenatal diabetes, 12 cancer, and 7 heart failure); audio recording of 28 face-to-face consultations (7 diabetes, 6 antenatal diabetes, 6 cancer, and 9 heart failure)	• What is said and done in consultations (video and face-to-face); unfolding interaction and strategies for communication; how technology is used in consultations (video and face-to-face); and how participants felt	• How people interact and communicate, how people create and maintain order and coherence in consultations together, and how video technology shapes, enables and constrains this; the relevance of different channels (verbal, visual, gesture, or gaze); and how these all shape the actions of users
Contextual data	• Accounts of 26 patients before/after the appointment (19 from VOCAL[a] and 7 from OTQS[b]) and 35 staff involved in delivering video consultations (28 from VOCAL and 7 from OTQS) combined with field notes from before/after face-to-face and video consultations at patient and clinician end • Documents (16 from VOCAL; 7 from OTQS) (eg, operating procedures and meeting minutes) • Researcher field notes about people and technologies delivering video consultations; including diagrams of how people, technologies, and clinical work interact • Demographic data	• Key interactions and interdependencies; key organizational routines and how these are changing over time; and accounts of clinical work and how this is shaped or reshaped through use of video consultations • Basic patient information, including age, gender, and ethnicity	• Social structures (eg, professional standards and definitions of excellence; what actors *know* and how they interpret the strategic terrain) and assumptions built into the technology about, for example, capability of users, how people interact, privacy and consent, the nature of clinical work, and how all these interact • Background and context to detailed micro-analysis

[a]VOCAL: Virtual Online Consultations-Advantages and Limitations.

[b]OTQS: Oxford Telehealth Qualitative Study.

Table 3. Overview of consultations in the Qualitative Analysis of Remote Consultations dataset.

Clinic	Total recorded	Male or female	Age (years), median (range)	Ethnicity
Diabetes (video)	12	5 male, 7 female	23 (21-50)	White British (5); white other (2); black Caribbean (1); Asian Bangladeshi (1); Asian Indian (3)
Diabetes (face-to-face)	6	3 male, 3 female	26 (21-58)	White British (2); black Caribbean (1); Asian Bangladeshi (2); Asian other (1)
Antenatal diabetes (video)	6	6 female	34 (30-37)	White British (1); black Caribbean (1); Asian Bangladeshi (1); Asian other (3)
Antenatal diabetes (face-to-face)	6	6 female	33 (26-36)	White British (0); black Caribbean (1); Asian Bangladeshi (3); Asian Indian (1); Asian other (1)
Cancer (video)	12	4 male, 8 female	74 (55-85)	White British (9); white other (1); black Caribbean (1); Asian Indian (1)
Cancer (face-to-face)	5	3 male, 2 female	69 (45-75)	White British (2); black Caribbean (2); Asian other (1)
Heart failure (video)	7	3 male, 4 female	67 (33-87)	White British (7)
Heart failure (face-to-face)	9	1 female, 8 male	60 (56-78)	White British (9)

Video consultations for the VOCAL study were recorded by a researcher (JW) using a small digital camcorder with a wide-angle lens. For the OTQS study, they were recorded by a researcher (CA/TG) using a small digital camcorder and a handheld iPad (Apple Inc). Face-to-face consultations were audio recorded using a digital Dictaphone. We captured clinician and patient interaction with the videoconferencing software using a commercially available screen-capture software tool (ACA Systems) to record screen images showing on each party's computer screen as a video file. This was run directly from a USB memory stick.

Each end of the consultation resulted in 2 digital files: 1 screen capture and 1 video. We synchronized these into 1 file using video editing software, meaning that the video of the computer screen could be played exactly in parallel with a video of the patient looking at the screen, and then aligned the patient and clinician *ends* (an example screenshot is shown in Multimedia Appendix 1). In several cases (Table 4), the consultation was recorded only at one end. This was either because the patient did not wish to be filmed in their home but was happy to be filmed from the clinic or because it was not practically possible to arrange for recording to take place at both ends.

We transcribed all face-to-face and video consultations using a specialized program, Transana [60], which allows researchers to capture the complexity of data, such as body language, while simultaneously following video, audio, and transcription (Multimedia Appendix 1).

In both studies, we made contemporaneous field notes at each patient's home and at the clinic, for instance, relating to the patient's material circumstances or the physical circumstances in which the clinician makes the remote call. We interviewed key staff at each site involved in setting up and conducting video consultations and patients as soon as possible after their video consultation.

Table 4. Summary of video consultation recordings.

Video recording	Cancer	Diabetes	Antenatal diabetes	Heart failure	Total
Dual clinic, dual home	4	0	1	2	7
Dual clinic, single home	1	0	1	3	5
Single clinic, dual home	1	2	1	1	5
Single clinic, single home	3	1	3	1	8
Dual clinic, no home	2	7	0	0	9
Single clinic, no home	1	2	0	0	3
Total	12	12	6	7	37

Theoretical Approach

We used 2 complementary theoretical approaches that see communication as a dynamic interaction that emerges moment by moment, which allowed us to examine the role of technology in shaping interaction. First, we used the *ethnography of communication* [61] to produce systematic and richly contextualized descriptions of the communicative genres, events, and practices that are observed in a particular culture to identify the key features of video consultations and attend systematically to contextual factors (eg, lack of spatial proximity and restricted visual field) that may produce differences with face-to-face consultations. Second, we used conversation analysis (CA) [62] to guide the fine-grained examination of the patterning of interaction, (ie, how consultations are managed by participants moment by moment). We focused on the linguistic, bodily, and contextual resources used by participants to create and maintain order and coherence and how different modes and channels of communication, such as the verbal and visual, shape actions (eg, physical examination).

Analysis

As is usual in linguistic ethnography, the analysis combined *zooming in* on interaction to understand how video consultations are successfully conducted, with *zooming out* to the clinic and wider organization to understand the context in which they take place. This process began with repeated viewings of video recordings among the team, which included representation from linguistics, sociology, psychology, and medicine with the aim of identifying "recurrent and stable details of talk-in-interaction" [62]. From this, we identified 4 questions for an in-depth study:

1. How do patients and clinicians open and close video consultations?
2. How are physical examinations accomplished in video consultations?
3. How do patients and clinicians address problems with the technology?
4. How is turn-taking affected by the video-mediated setting?

For the first 2 questions, we built collections of all openings, closings, and physical examinations, respectively. For the third question, we built a collection of cases in which the technology failed (eg, there was no audio) or where a failure in the technology was addressed (eg, a patient says his image is frozen). For the fourth question, we built a collection of cases where the patient and clinician struggled with smooth turn-taking, and so had repeated overlapping talk and/or stopped talking. Silences and overlapping talk are common in all conversations [54]; hence, not all cases of overlap and silence in our data were necessarily a result of the video medium. We therefore (1) only included consultations recorded at both patient and clinic ends, allowing us to determine that latency (ie, the transmission delay between one participant saying something and the other hearing it) was an issue, and (2) focused on points in the conversation where speaker transition should take place, but the latency caused problems.

In line with CA [63,64], we transcribed data using Jeffersonian and Mondada systems (Multimedia Appendix 2), employing a range of symbols and conventions to capture *what* participants say, *how* they say it (eg, pauses and volume), and accompanying *nonverbal* communication (eg, gaze and head movement). Guided by the principles of sequence organization and turn design [65,66], we then examined each case (eg, individual opening) to determine the actions that participants performed, how these actions were organized in the consultation [67,68], and the communication practices that participants used to make

these actions recognizable. We then searched for patterns and commonalities across each collection to distinguish between communication strategies that were relevant for single instances and those used more systematically, for instance, the ways in which clinicians use different opening questions. In other words, we identified the norms and conventions that participants appeared to rely on in video consultations.

Finally, once we had a solid appreciation of the patterns of communication across collections, we identified areas where there was *trouble*, such as potential misunderstandings, and grouped these into key challenges. At this point, we *zoomed out* to our contextual data with the aim of better understanding how and why such challenges occur (eg, the way in which Skype technology is set up on computers and what this requires of clinicians when they open video consultation) and how they are negotiated.

Patient and Public Involvement

Patients and their caregivers were key to the research. We set up a dedicated patient advisory group (PAG) in 2015 to enable patients to review documents and give feedback on their experiences and to incorporate patient feedback within the respective studies. We have since set up a standing group, Patients Active In Research on Digital health, who, along with the PAG, have reviewed emerging findings (including anonymized video clips) from the QuARC study and provided feedback on developing guidance on video consultations. Finally, we worked with a design company to co-design guidance and an accompanying animation that involved 2 workshops with a total of 50 members of the public, followed by a review of materials via email. This process was critical in developing both content and design of these outputs (eg, guiding us not only to focus on how to do a video consultation but also to include technical setup).

Results

Overview of Findings

We found that for this highly selective group of patients and clinicians, the use of video technology typically (but not always) enabled effective consultations that followed the conventional organization of a consultation [68]. Patients and clinicians adapted quickly and easily to the video-mediated context, transferring communication strategies from face-to-face communication to successfully complete web-based consultations. For example, in video consultations, the webcam is positioned above the screen, making mutual gaze impossible, but participants had no problem positioning their head and shoulders in front of the camera and looking at the screen to show that they were engaged and directing their attention to their coparticipant.

Clinicians managed to start consultations in the same way as a face-to-face consultation by asking, for instance, “How are you today?” and inviting patients to provide updates on their condition since the last consultation [69]. Patients responded in the conventional way, either reporting no problems or highlighting recent changes (eg, high blood sugars). When clinicians were engaged in other activities, such as updating the patient’s record, they made this clear by saying, for instance, “If I go silent, I’m writing. Okay?” Closing of consultations was straightforward, and followed the same pattern as face-to-face consultations, with the clinician checking if the patient had any further questions and then closing down the discussion, arranging a follow-up as needed, and saying goodbye.

The use of video calling (30 Skype, 7 FaceTime) was a novel form of communication for many users (particularly for older participants, Table 3). Younger patients appeared particularly comfortable with the technology, but even those with less technological affinity had few problems using the technology once they had some assistance, for instance, from a carer. When clinicians and users initially made contact, we observed that in many cases, consultations were characterized by surprise, amusement, and a sense of informality. This often continued into the consultation, with participants sometimes joking, generally, and about the use of video. It was striking that those clinicians who were new to the video medium felt quite distracted by having their own image appear on the screen for the first time—working out how to minimize or ignore it was an important step in re-establishing a formal, professional focus.

It was when we began to look at these issues in detail that we became interested in how it is that people interact effectively via video. Across the 37 video consultations in our dataset (12 hours and 21 min of total consultation time), we identified 3 key challenges that (some but not all) participants faced when doing video consultations: (1) establishing a connection and starting a video consultation, (2) dealing with disruption to conversational flow caused by technical issues with audio or video, and (3) conducting a physical examination via video. We describe each of these below, along with the communication strategies used to negotiate them. We present data extracts illustrating the analysis in the figures below (full transcriptions following CA conventions are included in Multimedia Appendix 3).

Challenge 1: Establishing a Connection and Starting a Video Consultation.

The start of a video consultation is a crucial time when the patient and clinician seek to establish a technical connection and determine if this is good enough for the consultation to go ahead. This process was not always straightforward. The combination of consultation and contextual data in our study enabled us to distinguish between *preopenings* involving steps to establish a connection and *openings* involving greetings like *hello*, followed by *how are you?* type questions [69] that signal the start of the clinical consultation.

Before a video consultation can get started, steps must be taken at both *ends* to enable the patient and clinician to establish a connection. For some participants, this opening sequence could be stressful, as they sought to ensure that the relevant equipment was in place and the technical connection was working to enable the clinical element of the consultation to begin. Depending on the technology, participants needed to first identify each other via their web-based user name (ie, exchanging remote addresses) and exchange phone numbers and/or email addresses (also, at the clinic end, ensure that internal procedures, standards,

software functionality, and administrative systems are in place to support the consultation; refer to Greenhalgh et al [70] for further details). The clinician would then usually *dial* the patient, who, by answering, would show that they were ready for the consultation. On the whole, preopenings ran smoothly. In some cases, participants experienced trouble, including outdated software that required *on the spot* updates (n=2), lost passwords (n=2), or misunderstandings about which platform to use (n=1). The problem was typically resolved by one participant (usually the clinician) calling the other by telephone, finding a solution, and then restarting the video connection.

Once a connection was established, opening greetings at the start of a video consultation were critical, not only in prompting conversation but also in indicating that participants could see and hear one another. It was at this point when technological issues, such as limited audio, were acknowledged and addressed. When clinicians and patients were familiar with the technology and potential problems, openings ran smoothly, and problems were quickly resolved. However, when one or more participants were not familiar with the technology, this could quickly lead to confusion and delay (often related to issues of breakdowns and latency, discussed below).

Take the example in Figure 1, which was taken from the start of a postoperative cancer consultation in which the patient's relative did not turn on the camera when she answered the doctor's video call; in the middle of line 1, when the daughter says *answer*, the doctor appears on their screen (screenshot, Figure 1).

Figure 1. Example of a clinician and patient establishing a connection at the start of a Skype consultation. Da: daughter; Dr: doctor; Ns: nurse.

```
01  Da:  oh. did I just answer wrong?
02  Dr:  (no)
03       (3.0)
04  Ns:  [(                )
05  Da:  [there's the doctor mum.
06  Dr:  (          )
07       (3.9)
08  Dr:  hello?
09       (.)
10  Da:  hi,
11       (0.4)
12  Ns:  hi. [(we) can't see you¿
13  Dr:      [hi.
14       (0.4)
15  Dr:  [we can hear you but we]
16  Da:  [ you  can't  see  us:?]
17  Dr:  can't see you; can you see
18       us?
19       (0.2)
20  Da:  I can see you:?
21       hold o:n,
22       (1.4)
23  Dr:  you need to (.) turn your
24       camera on?
25  Ns:  (oh)
```

In this example, the patient and clinician have already done the work of preopening, the patient has been invited to a video consultation, with Skype addresses and video appointment times confirmed to enable that to happen, and phone numbers confirmed ensuring an alternative means of communicating. Once a connection is established, the patient and clinician do not greet each other but instead query the connection and confirm at the patient's end that they can see the doctor. Then, 2 periods of silence follow (3.0 and 3.9 seconds, respectively). During this time, the doctor is unsure what is happening: by *not* greeting, the patient and her daughter do not provide the doctor with evidence that they have a working video connection meaning that, for the doctor and nurse at the other end, it is not clear if there is a problem with the connection or video. The doctor's *hello* with its strong rising pitch (line 8) is not a greeting, but a means of testing if the patient and/or daughter can hear him—a practice that we found repeatedly when there was a problem with the connection and participants wanted to see if they could still hear each other. When the daughter is heard to have responded with *hi* (line 10), this provides evidence that they can hear, that the connection works, and that the problem is purely visual (as the nurse points out in line 12). The daughter moves to turn the camera on, the doctor confirms that they can now see, and the consultation begins. Unlike in a face-to-face consultation, it is only at this point that the doctor asks *how are you*, signaling the start of the clinical consultation and enabling the social rituals of *being a doctor* and *being a patient* to begin.

Challenge 2: Dealing With Disruption to Conversational Flow

The technology used in video consultations is intended to facilitate interaction. In our dataset, there were only 10 (of 37) consultations in which participants experienced no problems

with the audio or video. In the other 27 consultations, we identified 49 instances in which a breakdown in audio and/or video disrupted the conversational flow. In 25 of these consultations, we had recordings at both *ends* (ie, clinician and patient; Table 3), allowing us to identify 151 cases in which latency disrupted interaction.

Such disruptions are significant as they not only disrupt conversational flow but also potentially impact the quality and outcome of consultations (Table 5). We provide a detailed analysis of the breakdowns and latency below.

Table 5. Frequency and duration of breakdowns and latency issues in video consultations.

Type of problem	Frequency[a]	Duration (range)
Audio or video breakdown at the start of a consultation		
No sound at the start		
Unsolved: clinician calls the patient on the phone using Skype for video	3	6.5 to 9.5 min
Solved: after disconnecting and reconnecting, the sound works	2	53 to 127 seconds
No video at the start		
Unsolved: camera does not work and participants make do with audio only	1	5 min and 5 seconds
Solved: either participant had forgotten to turn on the camera	8	6.8 to 22.3 seconds
Audio or video breakdown during a video consultation		
Garbled sound: the quality of the sound suddenly degrades, causing a problem with audibility, the consultation is halted, participants check when the audio works, and then resume the consultation	24	3.7 to 56.8 seconds
Bad audio throughout: the quality of the audio is poor because of technical problems, causing frequent noise or low volume	3	Continuous[b]
Video cutout: the video briefly cuts out on one side, before automatically resuming; may happen because of an incoming call	2	1.4 to 8.4 seconds
Bad video throughout: the quality of the video is poor because of a bad internet connection, causing the image to blur, freeze, or even cut out completely	1	Continuous[b]
Breakdown in connection during a video consultation		
Automatically solved: the connection cuts out briefly, but resumes automatically; participants briefly discuss and check if the connection works before resuming	2	6.1 to 15.4 seconds
Requiring reconnection: the connection is dropped completely and participants have to redial to get the connection back	3	43.5 to 71.9 seconds
Latency		
Brief overlap: participants talk at the same time, but either drops out after 1 or 2 syllables of overlapping talk	122	Up to 0.5 seconds
Competition for turn: participants talk in overlap for a while, using multiple explicit strategies to figure out whose turn it is	29	0.5 to 10.5 seconds

[a]Reporting the number of problems we identified in the dataset relating to latency and breakdowns.

[b]It is not possible to report exact duration as there were problems, either with audio or video, throughout. This results in continuous issues.

Breakdowns

Breakdowns in audio and/or video quality occurred at different times in video consultations and with variable duration and impact (Table 5). At the start of consultations, we identified 5 cases with no audio and 9 cases with no video. In the 5 cases with no audio, 3 were in the cancer service (because of technical problems caused by interference with other devices) and involved the clinician taking up to 9.5 min to try to solve the problem before calling the patient by phone and using video calling merely for the visual connection (once the other devices were identified, such interference was resolved), 1 involved the clinician disconnecting and reconnecting twice and using the inbuilt testing service (taking 2 min and 7 seconds), and 1 involved the patient disconnecting and then reconnecting (53 seconds). Audio breakdowns were only observed in Skype consultations.

Problems with the video connection at the start of consultations were typically the result of the patient or clinician not turning on their camera (8 cases). In 1 case, the clinician could not get the camera to work and, after just over 5 min, decided with the patient that they would proceed with an audio-only consultation.

Problems during consultations were often (but not always) disruptive. We found 24 cases where there was a minor problem with the audio resulting in, for instance, soft or garbled talk (Table 5). These lasted up to 57 seconds but were typically resolved in less than 10 seconds, with participants waiting for audio to be restored. The patient or clinician then pointed out that they had not heard something, and the other participant

then repeated their last sentence before continuing with the consultation.

In 4 consultations, the quality of the hardware or a slow internet connection caused persistent problems with audio (n=3) or video (n=1). In 2 cases, one with a low volume on the patient's end, the other with a repeated *frozen image* of the patient, participants attempted to solve the problem initially and then agreed to *make do*. In the other 2 cases, the audio frequently suffered from noise or distortions. For 1 consultation, this led to repeated minor problems (n=13), with each party repeatedly unable to hear, and halting the consultation to seek clarification. In the other, a postoperative cancer consultation, there was distortion making it hard to hear, resulting in a misunderstanding about the medication dose. In the extract in Figure 2, the doctor recommends that the patient shift to a higher dose formulation of Creon (pancreatic enzyme replacement)—25,000 instead of 10,000 units; however, the patient mishears and the level of disruption to the audio then limits opportunities for the patient to indicate that they have not understood, what conversation analysts refer to as limited *repair space* [55,71].

Figure 2. Example of significant disruption to a Skype consultation due to audio problems. Pt: patient.

```
01  Dr:  one suggestion,
02       (0.9)
03  Dr:  you're taking the ten thousand
04       capsules, [isn't it.
05  Pt:            [yes.
06       (.)
07  Dr:  they are also available at twenty-five
08       thousand.
09       (2.3)
10       so (0.4) so you could take (0.2) just
11       one capsule with light (.) meals,
12       and then two [or three for big meals.
13  Pt:               [ah.
14       I see.
15       (0.4)──────────────────────────────────── Start of audio
16       uhm at the (       )¿ they do not uhm     distortion
17       make (0.2) five (thousand) ones.
18       [uh (you) can only get ten thousand
19  Dr:  [hm
20  Pt:  ones. (   ) have to- yes.
21       I (could)- (.) I could (lower) the dose.
22       (.)
23       [yeah
24  Dr:  [you c- you could──────────────────────── End of distortion
25       ask your GP the next time you
26       [get the (creon),
27  Pt:  [mm hm,
```

In this example, the clinician's recommendation to change to a 25,000-unit strength capsule is indistinctly audible at the patient's end, but the patient appears to mishear and to understand that the consultant wants her to start taking 5000-unit strength capsules (line 17). She indicates that she understands with "ah I see" (lines 13-14), but then points out that Creon is not available in 5000-unit strength (lines 16-17), revealing her misunderstanding. The clinician then has the opportunity to rectify the problem [72], but audio distortion, caused at least in part by the microphone at the patient's end, means that the doctor appears to have trouble making out what the patient is saying and to assume that she heard correctly, recommending that the patient ask her GP for a change in unit strength (lines 24-26). The technical breakdown means that he does not notice that the patient misheard and so continues with the misunderstanding unnoticed and unresolved.

We found 2 cases in which the video was temporarily disrupted on one end of the consultation. In each case, the problem was quickly and automatically resolved (Table 5).

In 5 cases, the connection cut out completely: in 2, the connection was quickly and automatically restored, and in 3, participants had to reconnect themselves.

Latency

Spoken conversation is characterized by a turn-taking system, a set of rules participants use to determine whose turn it is to talk. This system serves to minimize both silence and overlapping talk in a conversation, the principle being 1 speaker at a time [54]. Research has shown that participants in a conversation favor direct responses (each person taking over immediately when the other has finished speaking) [73]. Delays in turn-taking and significant overlap usually indicate that there is a communication problem [74]. In our dataset, we found that short periods of latency (up to approximately 200 milliseconds) were tolerated or ignored, but that more significant latency

(approximately 500 milliseconds and above) interfered with this system, resulting in silence and overlapping talk or interruption.

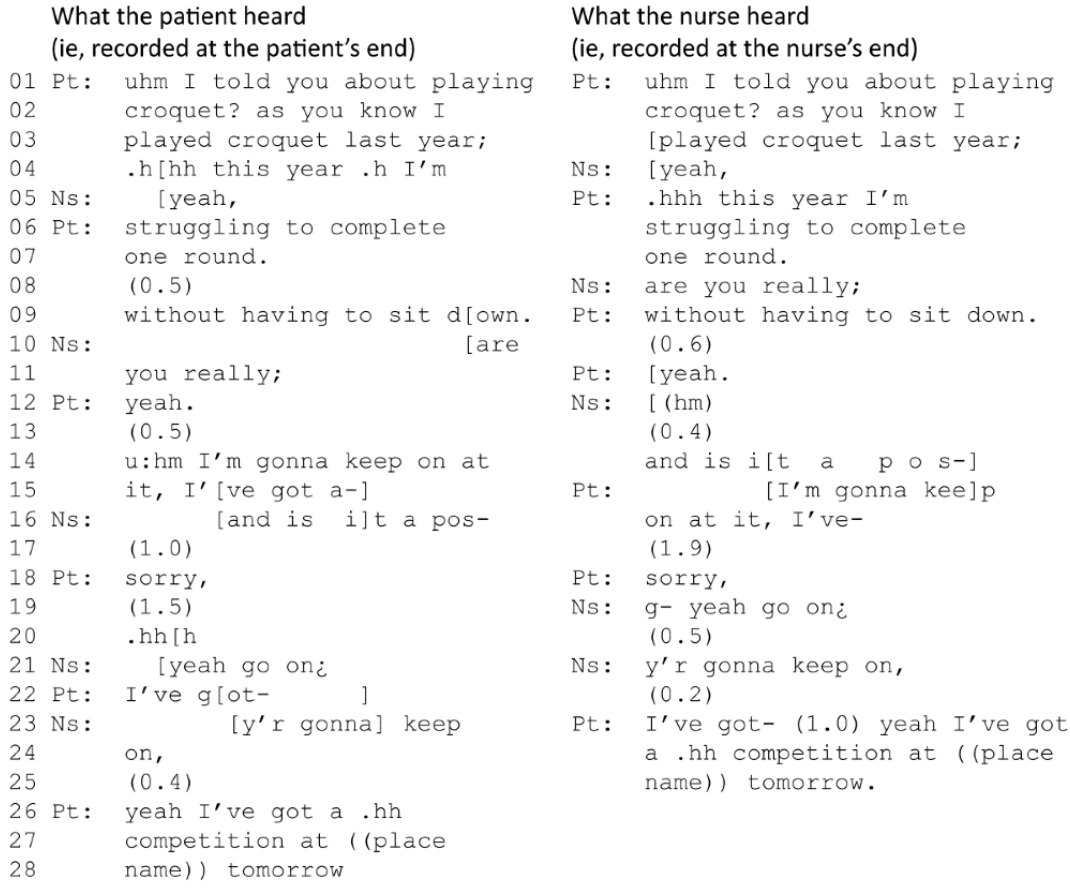
The 151 cases of significant latency that we found were unevenly distributed across our dataset, with 67 instances in 21 consultations in which problems were resolved in less than 10 seconds and another 84 instances in only 4 consultations in which problems were not only more frequent but typically took longer to resolve. In 1 of the 4 consultations, there were 47 instances of latency.

Of the 151 cases of latency, 122 were quickly resolved with 1 participant dropping out (ie, stopping talking) after 1 or 2 syllables of overlapping talk (Table 5), and the conversation quickly resumed with minimal impact on flow. In the remaining 29 cases (23 of which were from the 4 consultations with repeated instances of latency), problems were more disruptive and took longer to fix. Here, participants actively competed for the right to talk, resulting in complete sentences overlapping. Latency resulted in longer exchanges in which participants had to work out whose turn it was to talk, which regularly took between 0.5. 10.5 seconds. Conversations were resumed by participants saying something like *sorry* or *go on* and/or repeating the last thing the other person said.

Consider the example in Figure 3, from a heart failure consultation in which the patient is talking to a specialist heart failure nurse about their back problems. Latency in the connection makes the discussion challenging. As the patient tells the nurse about his back problem and the restrictions it places on his day-to-day life (column 1, line 15), the nurse seemingly interrupts to offer a solution (line 16) before the patient has resumed his story. This supposed *interruption* is a result of latency. As can be seen from the nurse's end (column 2), from her point of view, she does not begin to talk in the middle of the patient's turn (line 16), but actually starts her turn before the patient (line 15). Due to the latency in the call, she cannot know that the patient has already begun talking again; similarly, the patient cannot realize that the nurse has begun to offer a solution to his presented problem. The result is overlapping talk. Both also deal with different problems: from the patient's perspective, the nurse interrupts him; from the nurse's perspective, it is the other way around. However, the latency means that neither of them can know this. Eventually, both then stop talking, and it takes a series of silences and *repair sequences* (eg, the patient's *sorry*, and the nurse's *yeah go on*) before the patient can resume his narrative (line 22).

In the 29 cases where latency caused disruptions to conversational flow, participants relied on the same communication strategies they use in face-to-face conversation to try to negotiate the problem. However, latency also caused problems when using these strategies. For example, in Figure 3, the nurse tells the patient to *go on* (line 21), and the patient then resumes his talk as is usual when face-to-face but, because of latency in the connection, the nurse does not realize that the patient has resumed his talk. The nurse then repeats part of his talk (line 23) as an additional strategy to give him the turn, causing the patient to again stop talking (line 22) and confirm before definitively resuming his talk (line 26).

Figure 3. Example of latency disrupting conversational flow in a video consultation for heart failure. Ns: nurse; Pt: patient.



Challenge 3: Conducting Physical Examinations Via Video

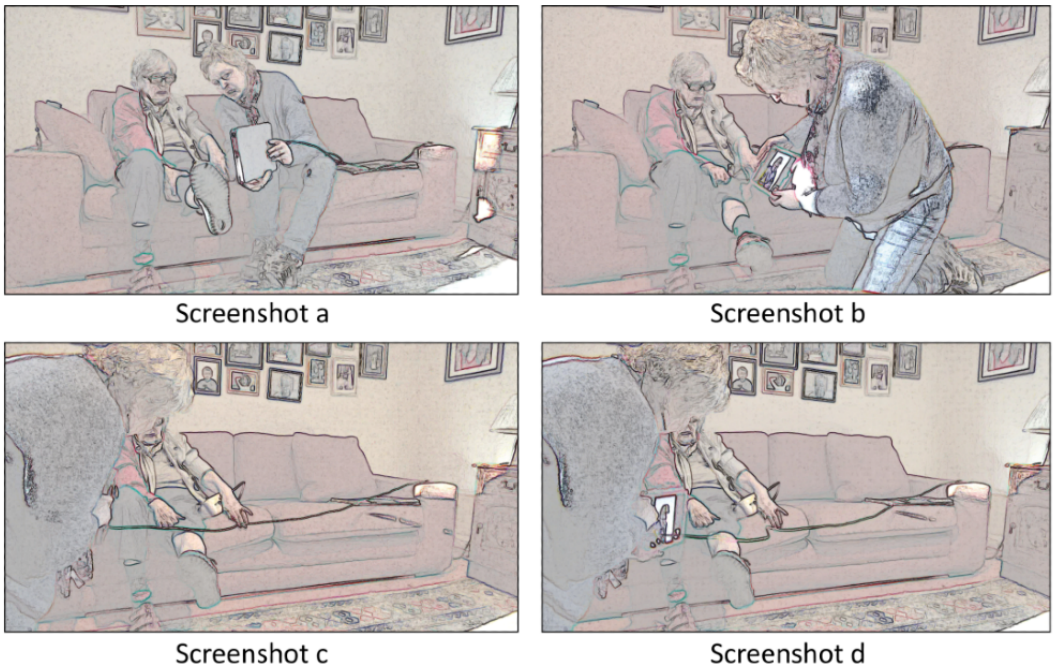
Clinicians and patients are concerned about the appropriateness and safety of conducting physical examinations in a video consultation [75,76]. Data from the heart failure service (including 7 video and 9 face-to-face consultations combined with contextual data about the setting; Table 2) provided a unique opportunity to examine what is possible in terms of physical examination in routine heart failure reviews between specialist heart failure nurses working from a community hospital and patients in their homes. Typically, in a face-to-face consultation, heart failure nurses measure weight, blood pressure, oxygen saturation, and heart rate and rhythm; they also assess edema (fluid build-up in the soft tissues), usually by examining the lower legs, and listening to the patient's chest with a stethoscope for signs of fluid overload or infection. With the exception of listening to the chest, the same examination can be attempted remotely, using a blood pressure monitor put on by the patient or relative (perhaps incorporating an irregular heartbeat indicator to estimate whether atrial fibrillation is present) and getting the patient or a relative to gently press on the lower leg (*digital pressure*) to produce an indentation (*pitting*) that is indicative of edema.

These examinations appeared to be very straightforward when the clinician and patient were colocated. However, in the remote setting, the same physical examinations presented specific challenges: (1) adequate design of instructions to guide video examinations (nurses had to explain tasks using lay language and check that instructions were being followed), (2) accommodation of the patient's desire for autonomy (on the part of nurses and relatives) in the light of opportunities for involvement in their own physical assessment, and (3) conducting a physical examination while simultaneously making it visible to the nurse (patients and relatives needed adequate technological knowledge to operate a device and make the examination visible to the nurse as well as basic biomedical knowledge to follow nurses' instructions). We have presented a detailed analysis of these challenges and the strategies used to attempt to overcome them in a separate paper [77].

Heart failure patients tend to be elderly, and many have poor exercise tolerance, mobility problems, or comorbidities, making attendance at the clinic potentially difficult. For the 9 patients in our sample attending the consultation in person, physical examinations were all successfully conducted by the clinician with minimal assistance from attending carers. For the 7 consulting by video, all participants (clinicians, patients, and, sometimes, relatives) had to employ novel communication strategies to collaboratively negotiate the challenges of achieving a satisfactory physical examination in the remote setting. Participants relied on a range of unspoken assumptions about the technology and their respective knowledge and understanding of medical procedures and the meaning of measurements. Clinicians and patients ran into unforeseen problems, either with the procedure (eg, the patient put a blood pressure meter incorrectly) and/or the technical process (eg, video quality prevented full visual assessment). Although complex, in 5 consultations, these problems were resolved with patients and/or their relatives asking the clinician for feedback or instructions and working with them to successfully complete the examination.

There were 2 consultations in which it was not possible to complete an attempted physical examination. In both instances, this was because the patient or relative was not able to hold the phone or tablet and test for edema at the same time. Take the example in Figure 4, of a heart failure patient and her daughter, in which they have been instructed by the clinician to examine the patient's right leg for edema; the daughter attempts to aim the camera at the patient's leg while at the same time trying to monitor the screen, as can be seen in the top 2 screenshots. This does not work adequately, and the nurse is not able to get a view of the patient. When she finally—after minutes of moving around—manages to aim the camera at the patient's leg (screenshot c, Figure 4), she is instructed by the nurse to press into the patient's leg. She lets go of the tablet with one hand and loses control (screenshot d, Figure 4), the result being that the camera is no longer aimed at the patient's leg. The nurse thus cannot assess if the patient has edema, and following this failed attempt, they agree to stop the examination. It is also worth noting that the patient end of the examination takes place in a living room that was not designed for clinical examinations. The sofa is much lower than a seat in a clinic, there is no examination couch, and the lighting in the home and on the tablet is not designed to provide the level of illumination considered standard for a clinical examination.

Figure 4. Heart failure patient and relative attempting examination for edema.



The 86-year-old patient shown in Figure 4 knows the nurse well and has seen her before in the clinic and at home. Having recently lost her husband, she had not been well and was admitted to the hospital with pneumonia. She deteriorated, resulting in close management by her GP and the specialist heart failure nurse. This story of acute deterioration in an older patient with multimorbidity is typical of the 16 patients with heart failure in our sample. Similar to the patient in Figure 4, those consulting via video were also doing so from a space (typically a living room) that was not designed for clinical examinations in which chairs were often too low, there was no examination couch, lighting was limited to standard bulbs, and there were often personal objects (eg, lamps or ornaments) that might impede movement or be knocked over. Hence, although 5 of the 7 remote examinations were completed, many found the process physically, practically, and technically challenging. Most were familiar with the technology from a clinic visit but had not necessarily practiced either in this kind of assessment or in using multiple devices to aid the process, especially in an environment that had not been designed with clinical assessment in mind. Unlike in face-to-face consultations, success was often dependent on the type of technology (smartphone, tablet, or laptop), the presence of a third party who could assist the patient, the patient's mobility, and the technological competence of the participants. Participants often reflected that it might be easier to perform a remote physical examination the second time as they were now aware of the problems they might encounter.

Discussion

Principal Findings

Our study has shown that, at least for this group of patients and clinicians, interaction in the majority of video consultations is effective. Observation and recording of video consultations, combined with theoretical sensitivity to the ways in which technology mediates communication in clinical consultations, has enabled us to do the following. First, we have shown that when participants experience technical or operational issues in video consultations (eg, when experiencing a problem connecting due to latency), they generally find a way to negotiate these, falling back on tried and tested rules of communication transferred from face-to-face situations. Familiarity with technology helps. Second, in a small proportion of consultations, technical interruptions before or during consultations may require workarounds such as using a standard telephone line to overcome lack of audio, and, in a few cases, it may threaten the quality of communication and clinical safety. Third, we have shown that in the minority of cases where significant latency occurs, clinicians and patients do not always hear the same thing at the same time. Finally, we have revealed that video consultations necessitate patients (and potentially carers too) to take an active role in understanding and manipulating the technology (eg, to change what the other party can see) and their surroundings (eg, adjusting lighting). For those consultations involving a remote physical examination, this involves a complex process of giving and receiving instructions and ensuring visual presence.

Strengths and Limitations

Careful analysis of verbal and nonverbal interaction, combined with a focus on social context, has enabled a robust understanding of the role played by language and social interaction in video consultations. Such detailed analysis takes time. Although we were fortunate to have data from 2 separate and fairly large qualitative studies, these comprised a relatively limited range of consultations in only 4 clinical conditions. Only 1 of these, heart failure, involved remote physical examinations.

CA is an established methodological technique for studying social interaction that allowed us to systematically identify challenges and to then generate and analyze relevant collections across our dataset. This is a key strength of our study. When the original studies were designed, we did not set out to examine

1 of these challenges—latency. This meant that the dataset we analyzed did not allow us to determine the precise timings of all instances of latency. When selecting data to analyze the impact of latency on conversational flow we therefore focused on cases where both participants talked in overlap following a point where turn transition could have taken place (eg, after a question) [54]. Future studies of latency in video consultations should ensure a comparison of recordings from both *ends* at every point.

Our dataset included video and audio recordings of consultations using Skype and FaceTime because these were the platforms being used at the time. Neither of these platforms was specifically designed for clinical consultations, and it may be that next-generation video consultation software (eg, Attend Anywhere) or more advanced peripheral technologies, such as noise-canceling microphones, may produce less in the way of breakdowns in conversations. We recommend that any future published studies of video interactions include details of the specific technologies used.

Comparison With Other Studies

To our knowledge, our study is one of the first to apply linguistic ethnographic approaches to the study of video-mediated consultations in health care. As such, it represents a significant addition to the existing literature, which has been dominated by trials and focused on issues of feasibility and acceptability and has also included some conventional CA without linguistic ethnography [44,46]. It extends our previous research on this dataset of video-mediated consultations, which include a multilevel analysis of the policy, organizational, and interactional aspects (using the Roter interaction analysis system) [9,47], and an in-depth analysis of physical examinations via video [77].

Our research confirms findings from our own and other studies that there are interactional differences between video and face-to-face consultations and potential for collaborative decision making via video [30,43,44,48,49,78]. Our study also corroborates that people tolerate some silence in conversation but work to minimize it [57] and that delays of 0.5 to 1 second can cause significant problems [52,53]. We have added to this, providing a detailed account of the frequency and duration of breakdown and latency issues specific to video consultations, showing that, in line with studies of how overlapping talk is negotiated in face-to-face conversations [79,80], participants have strategies for dealing with these problems.

Previous work has shown that speakers typically rely on what phenomenologists refer to as *reciprocity of perspectives* [81], meaning that both parties assume that the other hears and sees what they hear and see. Any differences, as far as speakers can be aware, are assumed to be irrelevant until proven otherwise [82]. Our study is the first to explore the relevance of this in video consultations in clinical settings, with 29 cases identified in which participants acted as if there were no delays and were not aware that there was a delay. This meant not only that it took them longer to solve the problem of overlapping talk but also that when they used conventional strategies for solving overlapping talk, this could lead to new problems (Figure 3). This has potentially significant implications for quality and safety, for instance, when communicating medication dosage.

Meaning of the Study

Our findings suggest that care is needed on the part of health care providers, commissioners, and policy makers in rolling out this new service model. Our own study and that of others has shown that video consultations appear to be largely safe and effective (albeit in a small sample of clinics and clinicians, and with patients identified by clinicians as *suitable* for a video consultation) [8,9,83], follow the conventional format of a face-to-face consultation, and any interactional challenges tend to be overcome. However, there is also a potential for significant problems, for instance, around miscommunication of dose or diagnosis, or misunderstandings because of technological breakdown.

Until recently, video consultations have tended to be an optional extra for many working in health services. The specter of COVID-19 has rapidly changed the landscape [84]. Clinicians and patients face a high risk of infection if they consult face-to-face, meaning that video consultations have significant advantages, both in cases where patients or clinicians are self-isolating because of symptoms of COVID-19 and where patients with other conditions are being seen remotely in an effort to reduce infection risks. In short, with a rapidly spreading disease, the pressure is on to reduce the number of people consulting face-to-face. Rolling out such a service at speed is, in some ways, similar to what we have studied in that many consultations are likely to involve dealing with routine, nonacute issues in a range of conditions with already known patients.

Before the COVID-19 pandemic, resources to support video consulting tended to be ad hoc and limited. In most cases, clinicians and patients were conducting video consultations without guidance for what works for which patients, clinical condition, or stage of diagnosis. That remains the case at the time of publication with, if anything, increased uncertainty about how video consultations can best be used as workflows are rapidly reorganized during the COVID-19 crisis. This lack of guidance means that clinicians and patients can run into unforeseen problems and communicative challenges that they may not know how to negotiate or easily resolve. Some problems potentially go unnoticed (eg, inadequate lighting misrepresents the patient's condition in a remote physical examination), which poses potential risks to the patient.

Guided by the findings from this and previous studies and in collaboration with patients (see above), we have developed a suite of freely available resources for patients and clinicians (available online [85] and summarized in Multimedia Appendix 4). This includes guidance on setting up and running video consultations for patients and clinicians, sets of frequently asked questions, downloadable leaflets and posters, and an animation. Our original intention, before the COVID-19 epidemic, was to provide a generic resource that could support patients or clinicians in setting up and running a video consultation. This remains the case. However, guided by emerging evidence, we have adapted guidance to ensure accessibility and relevance for those using video calling during and beyond the COVID-19 pandemic.

Unanswered Questions and Future Research

Communication in video consultations is under-researched. Further qualitative work is needed to examine interactional issues across a broader range of settings, conditions, and populations to inform the development of video consultations and to support patients and clinicians who choose to use them. This includes other kinds of remote physical examination for other conditions and with other types of patients (including those like older people or those with frailty who typically need more health care consultations but have limited experience of video technology).

Questions remain about the extent of latency, its often subtle effects [80], and the implications of how overlapping talk is understood differently by each participant. Further work is needed to consider a wider range of overlapping talk and to gain detailed understanding of how latency affects conversational flow and may cause misunderstandings in video consultations.

There have been significant advancements in technology in recent years, enabling both standalone platforms (eg, Skype) and those dedicated to video consultations (eg, Attend Anywhere). Future studies should examine what, if any, technical and interactional benefits the latter bring, including improved peripherals (eg, better screens or webcams) and use of companion devices (ie, multiple devices in combination).

Conclusions

Video consultations are interactionally different from face-to-face consultations. The use of video technology has the potential to change the way in which patients and clinicians interact, particularly when problems with audio or video interrupt the usual flow of interaction or where a physical examination is required. The use of evidence-based guidance, combined with training and support, can help clinicians and patients to more quickly identify and work through technical problems and avoid the potential for significant misunderstanding. Such resources are likely to be essential if video consultations are to be delivered at scale.

Acknowledgments

The authors thank the staff and patients who participated in the VOCAL and OTQS studies and members of the study steering groups, especially Mrs Jenni Bowley, who chaired these. The authors also thank Chrysanthi Papoutsi, who collected some of the data at the clinician end of consultations in the OTQS study. This study was funded by the NIHR Research for Patient Benefit program (grant reference number PB-PG-1216-20012). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. Development of guidance on doing a video consultation and the accompanying animation for patients was additionally supported by a Wellcome Trust Senior Investigator Award to TG (grant reference WT104830MA) and a Health Foundation grant (to support the extension of video consultations in England, scale-up improvement award to TG, SS, JW, and NHS partners). SS and TG are partly funded by the NIHR Biomedical Research Centre (BRC), Oxford, United Kingdom (grant reference number NIHR-BRC-1215-20008).

Authors' Contributions

SV had the original idea of introducing virtual consultations at Barts Health and established the service as a pilot in 2010. CA introduced video consultations at Oxford University Hospitals Trust, piloting the service in 2016. CA is a clinical academic and cardiovascular lead for the Oxfordshire Clinical Commissioning Group. SB and SV are consultants, and JM is a senior project manager at Barts Health. All were involved in refining study design, facilitating access to the study sites, data collection, and patient engagement in the respective OTQS (Oxford) and VOCAL (Barts Health) studies. TG was the chief investigator of the OTQS and VOCAL studies. SS had the initial idea for combining data from VOCAL and OTQS and was the chief investigator for this study and (as such) its guarantor. She was involved in all aspects of study design, data collection and analysis, and writing; she drafted the first version of the paper. JW is a senior academic involved in all aspects of the research. He led ethical applications for VOCAL and OTQS, on the meso- and microelements of the VOCAL study and on data collection for OTQS. DC is a senior academic with expertise in sociolinguistics, and LMS is a postdoctoral researcher specializing in CA. Both contributed methodological perspectives on the analysis of language and communication. LS led microanalysis of video consultation data. All authors have reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screenshot of Transana workspace.
[DOCX File , 289 KB - jmir_v22i5e18378_app1.docx]

Multimedia Appendix 2
Transcriptions conventions for conversation analysis.
[DOCX File , 15 KB - jmir_v22i5e18378_app2.docx]

Multimedia Appendix 3
Full transcriptions for Figures 1-3, using conversation analysis conventions.
[DOCX File , 690 KB - jmir_v22i5e18378_app3.docx]

Multimedia Appendix 4
Summary guidance on video consultations for clinicians and patients.
[DOCX File , 45 KB - jmir_v22i5e18378_app4.docx]

References

1. United Kingdom National Health Service. 2019. NHS Long Term Plan URL: https://www.longtermplan.nhs.uk/ [accessed 2020-04-29]
2. Honeyman M. The King's Fund. 2016. A Digital Nhs? An Introduction to the Digital Agenda and Plans for Implementation URL: https://www.kingsfund.org.uk/publications/digital-nhs [accessed 2020-04-29]
3. Wachter RM. Assets: Government of UK. 2016. Making IT Work: Harnessing the Power of Health Information Technology to Improve Care in England. Report of the National Advisory Group on Health Information Technology in England URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/550866/Wachter_Review_Accessible.pdf [accessed 2020-04-29]
4. de Oliveira C, Bremner KE, Ni A, Alibhai SM, Laporte A, Krahn MD. Patient time and out-of-pocket costs for long-term prostate cancer survivors in Ontario, Canada. J Cancer Surviv 2014 Mar;8(1):9-20. [doi: 10.1007/s11764-013-0305-7] [Medline: 23975612]
5. Dang S, Gomez-Orozco CA, van Zuilen MH, Levis S. Providing dementia consultations to veterans using clinical video telehealth: results from a clinical demonstration project. Telemed J E Health 2018 Mar;24(3):203-209. [doi: 10.1089/tmj.2017.0089] [Medline: 28686082]
6. Sechrist S, Lavoie S, Khong C, Dirlikov B, Shem K. Telemedicine using an iPad in the spinal cord injury population: a utility and patient satisfaction study. Spinal Cord Ser Cases 2018;4:71 [FREE Full text] [doi: 10.1038/s41394-018-0105-4] [Medline: 30131874]
7. Ignatowicz A, Atherton H, Bernstein CJ, Bryce C, Court R, Sturt J, et al. Internet videoconferencing for patient-clinician consultations in long-term conditions: a review of reviews and applications in line with guidelines and recommendations. Digit Health 2019;5:2055207619845831 [FREE Full text] [doi: 10.1177/2055207619845831] [Medline: 31069105]
8. Armfield NR, Bradford M, Bradford NK. The clinical use of Skype-for which patients, with which problems and in which settings? A snapshot review of the literature. Int J Med Inform 2015 Oct;84(10):737-742. [doi: 10.1016/j.ijmedinf.2015.06.006] [Medline: 26183642]
9. Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, Hanson P, et al. Advantages and limitations of virtual online consultations in a NHS acute trust: the VOCAL mixed-methods study. Health Serv Del Res 2018;6(21):1-136. [Medline: 29939519]
10. Shaw SE, Cameron D, Wherton J, Seuren LM, Vijayaraghavan S, Bhattacharya S, et al. Technology-enhanced consultations in diabetes, cancer, and heart failure: protocol for the qualitative analysis of remote consultations (QUARC) project. JMIR Res Protoc 2018 Jul 31;7(7):e10913 [FREE Full text] [doi: 10.2196/10913] [Medline: 30064972]
11. Freeman KA, Duke DC, Harris MA. Behavioral health care for adolescents with poorly controlled diabetes via Skype: does working alliance remain intact? J Diabetes Sci Technol 2013 May 1;7(3):727-735 [FREE Full text] [doi: 10.1177/193229681300700318] [Medline: 23759406]
12. Hansen CR, Perrild H, Koefoed BG, Zander M. Video consultations as add-on to standard care among patients with type 2 diabetes not responding to standard regimens: a randomized controlled trial. Eur J Endocrinol 2017 Jun;176(6):727-736. [doi: 10.1530/EJE-16-0811] [Medline: 28325823]
13. Harris MA, Freeman KA, Duke DC. Seeing is believing: using Skype to improve diabetes outcomes in youth. Diabetes Care 2015 Aug;38(8):1427-1434. [doi: 10.2337/dc14-2469] [Medline: 26033508]
14. Bertuzzi F, Stefani I, Rivolta B, Pintaudi B, Meneghini E, Luzi L, et al. Teleconsultation in type 1 diabetes mellitus (TELEDIABE). Acta Diabetol 2018 Feb;55(2):185-192. [doi: 10.1007/s00592-017-1084-9] [Medline: 29209814]
15. Robinson MD, Branham AR, Locklear A, Robertson S, Gridley T. Measuring satisfaction and usability of FaceTime for virtual visits in patients with uncontrolled diabetes. Telemed J E Health 2016 Feb;22(2):138-143. [doi: 10.1089/tmj.2014.0238] [Medline: 26295592]
16. Host BK, Turner AW, Muir J. Real-time teleophthalmology video consultation: an analysis of patient satisfaction in rural western Australia. Clin Exp Optom 2018 Jan;101(1):129-134 [FREE Full text] [doi: 10.1111/cxo.12535] [Medline: 28436157]
17. Viers BR, Lightner DJ, Rivera ME, Tollefson MK, Boorjian SA, Karnes RJ, et al. Efficiency, satisfaction, and costs for remote video visits following radical prostatectomy: a randomized controlled trial. Eur Urol 2015 Oct;68(4):729-735. [doi: 10.1016/j.eururo.2015.04.002] [Medline: 25900782]

18. Somers TJ, Kelleher SA, Westbrook KW, Kimmick GG, Shelby RA, Abernethy AP, et al. A small randomized controlled pilot trial comparing mobile and traditional pain coping skills training protocols for cancer patients with pain. Pain Res Treat 2016;2016:2473629 [FREE Full text] [doi: 10.1155/2016/2473629] [Medline: 27891252]
19. Katz IJ, Pirabhahar S, Williamson P, Raghunath V, Brennan F, O'Sullivan A, et al. iConnect CKD-virtual medical consulting: a web-based chronic kidney disease, hypertension and diabetes integrated care program. Nephrology (Carlton) 2018 Jul;23(7):646-652. [doi: 10.1111/nep.13070] [Medline: 28474361]
20. Newman SD, Toatley SL, Rodgers MD. Translating a spinal cord injury self-management intervention for online and telehealth delivery: a community-engaged research approach. J Spinal Cord Med 2019 Sep;42(5):595-605. [doi: 10.1080/10790268.2018.1518123] [Medline: 30199344]
21. Shem K, Sechrist SJ, Loomis E, Isaac L. SCiPad: effective implementation of telemedicine using iPads with individuals with spinal cord injuries, a case series. Front Med (Lausanne) 2017;4:58 [FREE Full text] [doi: 10.3389/fmed.2017.00058] [Medline: 28611986]
22. Nield M, Hoo GW. Real-time telehealth for COPD self-management using Skype. J Chron Obstruct Pulmon Dis 2012 Dec;9(6):611-619. [doi: 10.3109/15412555.2012.708067] [Medline: 22946768]
23. Mark DD, Ikehara C, Matsuura C, Hara K, Li D. Validating the impact of teaching pursed-lips breathing with Skype. J Hosp Palliat Nurs 2013;15(8):424-432. [doi: 10.1097/njh.0000000000000015]
24. Choi NG, Hegel MT, Marti N, Marinucci ML, Sirrianni L, Bruce ML. Telehealth problem-solving therapy for depressed low-income homebound older adults. Am J Geriatr Psychiatry 2014 Mar;22(3):263-271 [FREE Full text] [doi: 10.1097/JGP.0b013e318266b356] [Medline: 23567376]
25. Chavooshi B, Mohammadkhani P, Dolatshahi B. A randomized double-blind controlled trial comparing davanloo intensive short-term dynamic psychotherapy as internet-delivered vs treatment as usual for medically unexplained pain: a 6-month pilot study. Psychosomatics 2016;57(3):292-300. [doi: 10.1016/j.psym.2016.01.001] [Medline: 27017450]
26. Ptomey LT, Szabo AN, Willis EA, Greene JL, Danon JC, Washburn RA, et al. Remote exercise for adults with down syndrome. Transl J Am Coll Sports Med 2018 Apr 15;3(8):60-65 [FREE Full text] [Medline: 29930988]
27. Pennington L, Stamp E, Smith J, Kelly H, Parker N, Stockwell K, et al. Internet delivery of intensive speech and language therapy for children with cerebral palsy: a pilot randomised controlled trial. BMJ Open 2019 Jan 30;9(1):e024233 [FREE Full text] [doi: 10.1136/bmjopen-2018-024233] [Medline: 30705241]
28. Bennell KL, Nelligan R, Dobson F, Rini C, Keefe F, Kasza J, et al. Effectiveness of an internet-delivered exercise and pain-coping skills training intervention for persons with chronic knee pain: a randomized trial. Ann Intern Med 2017 Apr 4;166(7):453-462. [doi: 10.7326/M16-1714] [Medline: 28241215]
29. Müller KI, Alstadhaug KB, Bekkelund SI. Acceptability, feasibility, and cost of telemedicine for nonacute headaches: a randomized study comparing video and traditional consultations. J Med Internet Res 2016 May 30;18(5):e140 [FREE Full text] [doi: 10.2196/jmir.5221] [Medline: 27241876]
30. Ekberg S, Danby S, Theobald M, Fisher B, Wyeth P. Using physical objects with young children in 'face-to-face' and telehealth speech and language therapy. Disabil Rehabil 2019 Jul;41(14):1664-1675. [doi: 10.1080/09638288.2018.1448464] [Medline: 29566569]
31. Krout RE, Baker FA, Muhlberger R. Designing, piloting, and evaluating an on-line collaborative songwriting environment and protocol using Skype telecommunication technology: perceptions of music therapy student participants. Music Ther Perspect 2010 Jan 1;28(1):79-85 [FREE Full text] [doi: 10.1093/mtp/28.1.79]
32. Fairweather GC, Lincoln MA, Ramsden R. Speech-language pathology teletherapy in rural and remote educational settings: decreasing service inequities. Int J Speech Lang Pathol 2016 Dec;18(6):592-602. [doi: 10.3109/17549507.2016.1143973] [Medline: 27063692]
33. Gund A, Sjöqvist BA, Wigert H, Hentz E, Lindecrantz K, Bry K. A randomized controlled study about the use of ehealth in the home health care of premature infants. BMC Med Inform Decis Mak 2013 Feb 9;13:22 [FREE Full text] [doi: 10.1186/1472-6947-13-22] [Medline: 23394465]
34. Orlandoni P, Peladic NJ, Spazzafumo L, Venturini C, Cola C, Sparvoli D, et al. Utility of video consultation to improve the outcomes of home enteral nutrition in a population of frail older patients. Geriatr Gerontol Int 2016 Jun;16(6):762-767. [doi: 10.1111/ggi.12551] [Medline: 26171754]
35. Westra I, Niessen FB. Implementing real-time video consultation in plastic surgery. Aesthetic Plast Surg 2015 Oct;39(5):783-790. [doi: 10.1007/s00266-015-0526-4] [Medline: 26169952]
36. Abimbola S, Keelan S, Everett M, Casburn K, Mitchell M, Burchfield K, et al. The medium, the message and the measure: a theory-driven review on the value of telehealth as a patient-facing digital health innovation. Health Econ Rev 2019 Jul 3;9(1):21 [FREE Full text] [doi: 10.1186/s13561-019-0239-5] [Medline: 31270685]
37. Mitzner TL, Stuck R, Hartley JQ, Beer JM, Rogers WA. Acceptance of televideo technology by adults aging with a mobility impairment for health and wellness interventions. J Rehabil Assist Technol Eng 2017;4:2055668317692755 [FREE Full text] [doi: 10.1177/2055668317692755] [Medline: 31186923]
38. Mozer R, Bradford NK, Caffery LJ, Smith AC. Identifying perceived barriers to videoconferencing by rehabilitation medicine providers. J Telemed Telecare 2015 Dec;21(8):479-484. [doi: 10.1177/1357633X15607136] [Medline: 26556061]

39. Peterson CM, Apolzan JW, Wright C, Martin CK. Video chat technology to remotely quantify dietary, supplement and medication adherence in clinical trials. Br J Nutr 2016 Nov;116(9):1646-1655 [FREE Full text] [doi: 10.1017/S0007114516003524] [Medline: 27753427]
40. Berner J, Anderberg P, Rennemark M, Berglund J. Case management for frail older adults through tablet computers and Skype. Inform Health Soc Care 2016 Dec;41(4):405-416. [doi: 10.3109/17538157.2015.1033528] [Medline: 26691495]
41. Vincze L, Rollo ME, Hutchesson MJ, Callister R, Thompson DI, Collins CE. Postpartum women's perspectives of engaging with a dietitian and exercise physiologist via video consultations for weight management: a qualitative evaluation. Healthcare (Basel) 2018 Jan 19;6(1):e8 [FREE Full text] [doi: 10.3390/healthcare6010008] [Medline: 29351233]
42. Haig-Ferguson A, Loades M, Whittle C, Read R, Higson-Sweeney N, Beasant L, et al. 'It's not one size fits all'; the use of videoconferencing for delivering therapy in a specialist paediatric chronic fatigue service. Internet Interv 2019 Mar;15:43-51 [FREE Full text] [doi: 10.1016/j.invent.2018.12.003] [Medline: 30619719]
43. Ekberg S, Shaw A, Kessler D, Malpass A, Barnes R. Orienting to emotion in computer-mediated cognitive behavioral therapy. Res Lang Soc Interact 2016 Sep 26;49(4):310-324 [FREE Full text] [doi: 10.1080/08351813.2016.1199085]
44. Pappas Y, Seale C. The opening phase of telemedicine consultations: an analysis of interaction. Soc Sci Med 2009 Apr;68(7):1229-1237. [doi: 10.1016/j.socscimed.2009.01.011] [Medline: 19201514]
45. Pappas Y, Seale C. The physical examination in telecardiology and televascular consultations: a study using conversation analysis. Patient Educ Couns 2010 Oct;81(1):113-118. [doi: 10.1016/j.pec.2010.01.005] [Medline: 20144523]
46. Pappas Y, Vseteckova J, Mastellos N, Greenfield G, Randhawa G. Diagnosis and decision-making in telemedicine. J Patient Exp 2019 Dec;6(4):296-304 [FREE Full text] [doi: 10.1177/2374373518803617] [Medline: 31853485]
47. Greenhalgh T, Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, et al. Real-world implementation of video outpatient consultations at macro, meso, and micro levels: mixed-method study. J Med Internet Res 2018 Apr 17;20(4):e150 [FREE Full text] [doi: 10.2196/jmir.9897] [Medline: 29625956]
48. Stommel W, van Goor H, Stommel M. Other-attentiveness in video consultation openings: a conversation analysis of video-mediated versus face-to-face consultations. J Comput Mediat Commun 2019;24(6):275-292. [doi: 10.1093/jcmc/zmz015]
49. Stommel W, Stommel M. Participation of companions in video-mediated medical consultations: a micro-analysis. In: Giles D, Meredith J, Stommel W, editors. Online Social Interaction. Basingstoke, England: Palgrave Macmillan; 2020.
50. Mlynář J, González-Martínez E, Lalanne D. Situated organization of video-mediated interaction: a review of ethnomethodological and conversation analytic studies. Interact Comput 2018;30(2):73-84. [doi: 10.1093/iwc/iwx019]
51. Heath C, Luff P. Disembodied conduct: interactional asymmetries in video mediated communication. In: Button G, editor. Technology in Working Order: Studies of Work, Interaction, and Technology. New York, USA: Routledge; 1992:35-54.
52. Ruhleder K, Jordan B. Co-constructing non-mutual realities: delay-generated trouble in distributed interaction. Comput Supported Coop Work 2001 Mar;10(1):113-138 [FREE Full text] [doi: 10.1023/a:1011243905593]
53. Isaacs E, Morris T, Rodriguez T, Tang J. A Comparison of Face-to-Face and Distributed Presentations. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 1995 Presented at: CHI'95; March 15-17, 1995; Maryland, United States. [doi: 10.1145/223904.223950]
54. Sacks H, Schegloff EA, Jefferson G. A simplest systematics for the organization of turn-taking for conversation. Language 1974 Dec;50(4):696. [doi: 10.2307/412243]
55. Schegloff EA, Jefferson G, Sacks H. The preference for self-correction in the organization of repair in conversation. Language 1977 Jun;53(2):361. [doi: 10.2307/413107]
56. Rusk F, Pörn M. Delay in L2 interaction in video-mediated environments in the context of virtual tandem language learning. Linguist Educ 2019 Apr;50:56-70. [doi: 10.1016/j.linged.2019.02.003]
57. Kendrick KH, Torreira F. The timing and construction of preference: a quantitative study. Discourse Process 2014 Aug 30;52(4):255-289. [doi: 10.1080/0163853x.2014.955997]
58. Rintel S. Omnirelevance in technologised interaction: couples coping with video calling distortions. In: Fitzgerald R, Housley W, editors. Advances in Membership Categorisation Analysis. Thousand Oaks, CA: Sage Publications; 2015:123-150.
59. Morris J, Campbell-Richards D, Wherton J, Sudra R, Vijayaraghavan S, Greenhalgh T, et al. Webcam consultations for diabetes: findings from four years of experience in Newham. Pract Diab 2017 Mar 20;34(2):45-50. [doi: 10.1002/pdi.2078]
60. Woods D. Transana v3.32. Madison,Wisconsin; 2020. URL: https://www.transana.com/ [accessed 2020-04-30]
61. Gumperz JJ, Hymes D, editors. Directions in Sociolinguistics: The Ethnography of Communication. Second Edition. Oxford, UK: Basil Blackwell; 1991.
62. Sidnell J, Stivers T, editors. The Handbook of Conversation Analysis. Chichester, UK: Wiley-Blackwell; 2012.
63. Heritage J, Maynard DW, editors. Communication in Medical Care: Interaction Between Primary Care Physicians and Patient. Cambridge, USA: Cambridge University Press; 2006.
64. Toerien MC. Conversations and conversation analysis. In: Flick U, editor. The Sage Handbook of Qualitative Data Analysis. Thousand Oaks, CA: Sage Publications; 2014:327-340.
65. Schegloff EA, editor. Sequence Organization in Interaction: A Primer in Conversation Analysis Vol I. Cambridge, UK: Cambridge University Press; 2007.

66. Drew PT. Turn design. In: Sidnell J, Stivers T, editors. The Handbook of Conversation Analysis. Chichester, UK: Wiley-Blackwell; 2013:131-149.
67. Robinson JD. An interactional structure of medical activities during acute visits and its implications for patients' participation. Health Commun 2003;15(1):27-57. [doi: 10.1207/S15327027HC1501_2] [Medline: 12553776]
68. Robinson JD. Overall structural organization. In: Sidnell J, Stivers T, editors. The Handbook of Conversation Analysis. Cambridge, UK: Wiley-Blackwell; 2013:257-279.
69. Heritage J, Robinson JD. The structure of patients' presenting concerns: physicians' opening questions. Health Commun 2006;19(2):89-102. [doi: 10.1207/s15327027hc1902_1] [Medline: 16548700]
70. Greenhalgh T, Wherton J, Shaw S, Papoutsi C, Vijayaraghavan S, Stones R. Infrastructure revisited: an ethnographic case study of how health information infrastructure shapes and constrains technological innovation. J Med Internet Res 2019 Dec 19;21(12):e16093 [FREE Full text] [doi: 10.2196/16093] [Medline: 31855184]
71. Robinson JD. What 'what?' tells us about how conversationalists manage intersubjectivity. Res Lang Soc Interact 2014 May 12;47(2):109-129. [doi: 10.1080/08351813.2014.900214]
72. Schegloff EA. Repair after next turn: the last structurally provided defense of intersubjectivity in conversation. Am J Sociol 1992 Mar;97(5):1295-1345. [doi: 10.1086/229903]
73. Sacks H. On the preference for agreement and contiguity in sequences in conversation. In: Button G, Lee JR, editors. Talk and Social Organisation. Clevedon, England: Multilingual Matters; 1987:54-69.
74. Pomerantz A. Agreeing and disagreeing with assessments: some features of preferred/dispreferred turn shapes. In: Atkinson JM, editor. Structures of Social Action: Studies in Conversation Analysis. Cambridge, UK: Cambridge University Press; 1984:57-101.
75. Fatehi F, Martin-Khan M, Smith AC, Russell AW, Gray LC. Patient satisfaction with video teleconsultation in a virtual diabetes outreach clinic. Diabetes Technol Ther 2015 Jan;17(1):43-48. [doi: 10.1089/dia.2014.0159] [Medline: 25296189]
76. Randhawa RS, Chandan JS, Thomas T, Singh S. An exploration of the attitudes and views of general practitioners on the use of video consultations in a primary healthcare setting: a qualitative pilot study. Prim Health Care Res Dev 2019 Jan;20:e5 [FREE Full text] [doi: 10.1017/S1463423618000361] [Medline: 29909798]
77. Seuren LM, Wherton J, Greenhalgh T, Cameron D, A'Court C, Shaw SE. Physical examinations via video for patients with heart failure: qualitative study using conversation analysis. J Med Internet Res 2020 Feb 20;22(2):e16694 [FREE Full text] [doi: 10.2196/16694] [Medline: 32130133]
78. Lee PA, Greenfield G, Pappas Y. The impact of telehealth remote patient monitoring on glycemic control in type 2 diabetes: a systematic review and meta-analysis of systematic reviews of randomised controlled trials. BMC Health Serv Res 2018 Jun 26;18(1):495 [FREE Full text] [doi: 10.1186/s12913-018-3274-8] [Medline: 29940936]
79. Schegloff EA. Overlapping talk and the organization of turn-taking for conversation. Lang Soc 2000 Jan 1;29(1):1-63 [FREE Full text] [doi: 10.1017/s0047404500001019]
80. Drew P. 'Quit talking while I'm interrupting': a comparison between positions of overlap onset in conversation. In: Haakana M, Laakso M, Lindstrom J, editors. Talk in Interaction: Comparative Dimensions. Helsinki, Finland: Finnish Literature Society; 2009:70-93.
81. Schutz A. Collected Papers I: The Problem of Social Reality. The Hague, The Netherlands: Martinus Nijhoff; 1962.
82. Heath C, Luff P. Technology in Action. Cambridge, UK: Cambridge University Press; 2000.
83. Fatehi F, Armfield NR, Dimitrijevic M, Gray LC. Clinical applications of videoconferencing: a scoping review of the literature for the period 2002-2012. J Telemed Telecare 2014 Oct;20(7):377-383. [doi: 10.1177/1357633X14552385] [Medline: 25399998]
84. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for covid-19: an opportunity in a crisis? Br Med J 2020 Mar 12;368:m998. [doi: 10.1136/bmj.m998] [Medline: 32165352]
85. Design Science. URL: https://design-science.org.uk/nhs-video-consulting/ [accessed 2020-04-30]

Abbreviations

BRC: Biomedical Research Centre
CA: conversation analysis
COVID-19: coronavirus disease 2019
GP: general practitioner
NHS: National Health Service
NIHR: National Institute for Health Research
OTQS: Oxford Telehealth Qualitative Study
PAG: patient advisory group
QuARC: Qualitative Analysis of Remote Consultations
VOCAL: Virtual Online Consultations-Advantages and Limitations

Edited by G Eysenbach; submitted 02.03.20; peer-reviewed by J Sturt; comments to author 30.03.20; revised version received 02.04.20; accepted 03.04.20; published 11.05.20.

Please cite as:
Shaw SE, Seuren LM, Wherton J, Cameron D, A'Court C, Vijayaraghavan S, Morris J, Bhattacharya S, Greenhalgh T
Video Consultations Between Patients and Clinicians in Diabetes, Cancer, and Heart Failure Services: Linguistic Ethnographic Study of Video-Mediated Interaction
J Med Internet Res 2020;22(5):e18378
URL: http://www.jmir.org/2020/5/e18378/
doi:10.2196/18378
PMID:32391799

Original Paper

Digital Care for Chronic Musculoskeletal Pain: 10,000 Participant Longitudinal Cohort Study

Jeannie F Bailey[1*], PhD; Vibhu Agarwal[2*], PhD; Patricia Zheng[1], MD; Matthew Smuck[3], MD; Michael Fredericson[3], MD; David J Kennedy[4], MD; Jeffrey Krauss[2,3], MD

[1]Department of Orthopaedic Surgery, University of California, San Francisco, San Francisco, CA, United States

[2]Hinge Health, Inc, San Francisco, CA, United States

[3]Division of Physical Medicine & Rehabilitation, Stanford University, Palo Alto, CA, United States

[4]Vanderbilt University Medical Center, Nashville, CA, United States

[*]these authors contributed equally

Corresponding Author:
Jeffrey Krauss, MD
Hinge Health, Inc
465 California St
14th Floor
San Francisco, CA, 94104
United States
Phone: 1 628 208 0776
Email: jeff.krauss@hingehealth.com

Abstract

Background: Chronic musculoskeletal pain has a vast global prevalence and economic burden. Conservative therapies are universally recommended but require patient engagement and self-management to be effective.

Objective: This study aimed to evaluate the efficacy of a 12-week digital care program (DCP) in a large population of patients with chronic knee and back pain.

Methods: A longitudinal observational study was conducted using a remote DCP available through a mobile app. Subjects participated in a 12-week multimodal DCP incorporating education, sensor-guided exercise therapy (ET), and behavioral health support with 1-on-1 remote health coaching. The primary outcome was pain measured by the visual analog scale (VAS). Secondary measures included engagement levels, program completion, program satisfaction, condition-specific pain measures, depression, anxiety, and work productivity.

Results: A total of 10,264 adults with either knee (n=3796) or low back (n=6468) pain for at least three months were included in the study. Participants experienced a 68.45% average improvement in VAS pain between baseline intake and 12 weeks. In all, 73.04% (7497/10,264) participants completed the DCP into the final month. In total, 78.60% (5893/7497) of program completers (7144/10,264, 69.60% of all participants) achieved minimally important change in pain. Furthermore, the number of ET sessions and coaching interactions were both positively associated with improvement in pain, suggesting that the amount of engagement influenced outcomes. Secondary outcomes included a 57.9% and 58.3% decrease in depression and anxiety scores, respectively, and 61.5% improvement in work productivity. Finally, 3 distinct clusters of pain response trajectories were identified, which could be predicted with a mean 76% accuracy using baseline measures.

Conclusions: These results support the efficacy and scalability of a DCP for chronic low back and knee pain in a large, diverse, real-world population. Participants demonstrated high completion and engagement rates and a significant positive relationship between engagement and pain reduction was identified, a finding that has not been previously demonstrated in a DCP. Furthermore, the large sample size allowed for the identification of distinct pain response subgroups, which may prove beneficial in predicting recovery and tailoring future interventions. This is the first longitudinal digital health study to analyze pain outcomes in a sample of this magnitude, and it supports the prospect for DCPs to serve the overwhelming number of musculoskeletal pain sufferers worldwide.

(J Med Internet Res 2020;22(5):e18250) doi:10.2196/18250

KEYWORDS

musculoskeletal pain; low back pain; patient engagement; exercise therapy; telemedicine; telerehabilitation; mobile phone

Introduction

Background

Chronic musculoskeletal pain has vast global prevalence [1] and annual costs in the hundreds of billions of dollars in the United States [2,3]. Musculoskeletal disorders are debilitating and may contribute to the opioid epidemic, as they are the most common noncancer indication for an opioid prescription in the United States [4-6]. Nonsurgical care, including exercise, education, and behavioral health, is universally recommended as the first-line treatment for the majority of chronic musculoskeletal conditions [7] given that it can achieve similar outcomes to surgery with reduced cost and lower risk [8,9]. However, conservative care has significant barriers to effective implementation and requires higher patient engagement to be successful [10,11]. Notably, conservative care administered in a clinical setting is also costly, and ongoing monitoring is often infeasible. Given the growing burden of chronic musculoskeletal pain, a scalable and effective mode of conservative care delivery is needed.

Digital health interventions have the potential to improve conservative care outcomes for chronic musculoskeletal pain by increasing patient engagement through electronic delivery of interventions. This approach can better enable patients to take a proactive role in their treatment and learn to self-manage their chronic pain symptoms. With the ubiquity of smartphones, low-cost sensor technology, and advanced analytical approaches to assess complex health care data, the prospect of digital technology for improved patient care is apparent and is reflected in the growing number of clinical trial protocols and review papers on the topic [12]. Digital therapies are shown to be effective for improving outcomes associated with conditions requiring self-management and behavioral change, such as type 2 diabetes [13], hypertension [14], and insomnia [15]. In addition, patient willingness to seek surgical treatment is shown to decrease following participation in a digital care program (DCP) [16]. Chronic pain, although often difficult to diagnose and treat clinically, is also shown to improve with the aid of digital therapy [10]. For chronic musculoskeletal pain specifically, the DCP in this study was previously evaluated in two randomized control trials and demonstrated effectiveness for improving pain and disability associated with knee pain [17] and low back pain [18]. Although these previous musculoskeletal pain studies show potential for a digital therapeutic approach to improve outcomes, they are limited in sample size (<200 subjects) and real-world effectiveness has yet to be shown. In this study, we assessed engagement and subject-reported outcomes over a 12-week period following enrollment in the DCP in a sample of over 10,000 users with chronic knee or back pain.

Objectives

This study had two objectives. First, we sought to determine whether the DCP is scalable and effective in a large sample of real-world patients. Given the magnitude of the chronic musculoskeletal pain population, scalability is one of the greatest potential benefits of a DCP, so the efficacy of a DCP in a large sample of real-world patients is important to assess. Key questions include if high levels of engagement can be sustained and if efficacy demonstrated in smaller randomized control trials is maintained in the larger real-world population. On the basis of results from the smaller randomized control trials, we hypothesized that the DCP would improve subject-reported pain over a 12-week period and that engagement with the DCP would be a necessary factor for improvement. A scalable digital intervention for engaging patients with safe conservative therapies for lasting self-management would have the potential to reduce the economic burden and improve the quality of life for a large population of patients.

Second, we sought to analyze the large dataset generated from the DCP to generate novel insights into patient recovery trajectories, which would create an opportunity to develop personalized interventions for individual patients. Little is known about the patient-specific response and rate of improvement for chronic musculoskeletal pain between clinical visits. Patients are typically assessed by clinicians during initial evaluations and, then, at follow-up appointments that may be weeks or months apart. A DCP enables regular (eg, weekly) collection of subject-reported outcomes throughout the recovery process. Statistical modeling methods can then be applied to these large longitudinal datasets to assess the rate of change in outcomes and if baseline data can predict recovery response. In this study, we used statistical modeling on a large longitudinal sample to evaluate nonlinear changes in pain over time and predict subject-specific pain response groups (rapid vs gradual) from baseline demographic data. Understanding how pain improves over time would inform our knowledge of pain recovery, identify variables associated with recovery, and allow for better care of patients unlikely to have rapid pain responses.

Methods

Study Design

This was a retrospective cohort study of consecutively recruited participants. Employees and their dependents at 30 participating employers across the United States were invited to complete a web-based application to participate in the Hinge Health DCP. Employees were diverse and included both office and service-based roles such as data analysts, manual laborers, truck drivers, catering staff, and outdoor instructors. Participants with low back or knee pain were recruited through email, direct mail, and posters. The trial was approved by the Western Institutional Review Board and complied with all ethical regulations. Participants provided informed consent and completed the intervention remotely. Each participant participated in 1 of 2 digital care pathways: 1 for chronic knee pain and the other for chronic low back pain. The only differences between the 2 pathways were the specific exercise regimens and some condition-specific education materials (eg, anatomy and surgical options). To mitigate the risks of selection bias, we included all

participants who had registered in the Hinge Health program by the cutoff date (May 6, 2019). We were able to verify that the study sample provided adequate power (after correcting for intrauser clustering effects, a sample size of 10,000 gave us a power of 0.97 to detect a 5-point change in our primary outcome with a type 1 error rate of 0.01). A summary of the key attributes of the cohort is provided in Table 1.

Inclusion criteria to qualify for participation in the DCP included being ≥18 years and not >80 years at the time of enrollment, having at least 12 weeks of back or knee pain, and having a baseline visual analog scale (VAS) score for pain greater than 0. Additional inclusion criteria for this study included starting the DCP, defined as completing at least one exercise session or reading 1 educational paper in the first 2 weeks following registration. Participants were excluded during registration by completing a screening questionnaire, which rejected patients with *red flag* symptoms, including signs of fracture, joint instability, infection, cancer, and cauda equina syndrome. Thus, this study included all consecutively qualified participants who enrolled in the DCP between February 6, 2017, and May 6, 2019, meeting the above inclusion and exclusion criteria

Table 1. Demographics and outcome measures (N=10,264).

Variables	Baseline			Final		
	Overall	Back pain (n=6468)	Knee pain (n=3796)	Overall	Back pain (n=6468)	Knee pain (n=3796)
Age (years), mean (SD)	43.57 (11.14)	42.58 (10.91)	45.26 (11.33)	N/A[a]	N/A	N/A
BMI, mean (SD)	30.25 (7.42)	29.76 (7.11)	31.09 (7.84)	N/A	N/A	N/A
Gender						
Female, n (%)	5132 (50.00)	4981 (48.53)	5388 (52.49)	N/A	N/A	N/A
Measures, mean (SD)						
Pain (VAS[b])	45.13 (22.42)	45.81 (22.16)	43.98 (22.81)	14.24 (15.31)	14.23 (15.12)	14.33 (15.59)
PHQ-9[c]	3.05 (5.34)	3.35 (5.49)	2.54 (5.04)	1.85 (3.97)	2.12 (4.12)	1.43 (3.38)
PHQ-9[d]	12.01 (4.61)	11.99 (4.56)	12.06 (4.73)	5.05 (5.72)	5.10 (5.73)	4.95 (5.70)
GAD-7[e]	3.93 (5.50)	4.39 (5.69)	3.15 (5.08)	2.21 (3.83)	2.48 (3.99)	1.77 (3.51)
GAD-7[f]	11.49 (4.10)	11.56 (4.13)	11.32 (4.04)	4.78 (5.05)	4.84 (5.01)	4.65 (5.12)
One-year surgery likelihood (0-100)	12.67 (21.55)	9.07 (17.89)	18.80 (25.51)	4.14 (12.44)	2.88 (9.26)	6.26 (16.1)
WPAI[g] (0-100)	31.74 (26.79)	34.12 (26.37)	27.54 (27.02)	11.45 (15.60)	12.24 (15.58)	10.17 (15.57)
KOOS—pain[h]	N/A	N/A	15.23 (6.66)	N/A	N/A	10.04 (5.81)
Modified von Korff	N/A	15.95 (5.03)	N/A	N/A	7.75 (5.44)	N/A

[a]N/A: not applicable.

[b]VAS: visual analog scale.

[c]PHQ-9: patient health questionnaire 9-item scale.

[d]The mean and SD of the scores in depressed (PHQ-9>5) subjects.

[e]GAD-7: generalized anxiety disorder 7-item scale.

[f]The mean and SD of the scores in anxious (GAD-7>5) subjects.

[g]WPAI: work productivity and activity impairment.

[h]KOOS—pain: knee injury and osteoarthritis outcome score—pain subscale.

Digital Care Program

Following registration, participants received a tablet computer via mail with the Hinge Health app installed, along with 2 Bluetooth wearable motion sensors with straps and instructions to be placed above and below the painful region during the in-app exercise therapy (ET). In the lower back program, a sensor was placed on the posterior lower back and anterior chest, and for the knee program, a sensor was placed over the anterior tibia and thigh. Sensors utilized standard accelerometer and gyrometer technology (InvenSense MPU-6050, TDK Electronics, Tokyo, Japan) and were used to objectively monitor compliance and performance of exercises. ET sessions comprised light-intensity stretching and strengthening exercises commonly used in clinical practice. The ET sessions were administered using animations and instructional videos to demonstrate how to perform each exercise. While performing the exercise, the app then displayed real-time graphics showing the position of the user's relevant body parts based on the wearable sensors and indicated if the exercise was within the

desired range of movement (see Multimedia Appendices 1 and 2.

Participants were assigned a personal coach and communication was performed via text message, email, or in-app messaging throughout the DCP. Health coaches completed certification through a coaching school approved by the National Board for Health & Wellness Coaching. Coaches attempted to interact with participants via their preferred communication method at least weekly. Phone calls with the coach were also offered to participants up to 3 times during the DCP. Each participant was also placed on a peer support team of 20-30 participants that utilized a discussion forum within the app, as previous qualitative research showed this to be an important feature [19]. All app participation was completed remotely, at times and places chosen by the participant. Each week, participants were instructed to complete at least three sessions of sensor-guided ET, read 2 education papers, and log their symptoms at least twice. Participants were able to complete more ET sessions or read more education papers if desired. Behavior change topics were addressed through education papers and brief interactive modules, and focused on common cognitive behavioral therapy topics, including catastrophizing, active coping methods, and fear avoidance. Additional behavior change mechanisms used in the program included goal setting and tracking. Finally, participants were encouraged to engage in 3 aerobic exercise activities per week and perform up to 4 brief modules based on cognitive behavioral therapy between weeks 3 and 9. Each participant also maintained access to treatment as usual. The app was developed, owned, and sponsored by Hinge Health, Inc.

Outcomes

The primary outcome was VAS pain for the question "Over the past 24 hours, how bad was your [back/knee] pain?" from 0 (*none*) to 100 (*worst imaginable*). This was asked weekly during the 12-week period immediately after an ET session, and participants also had the option to report VAS unprompted, for a total of up to 2 pain scores per week. Our definition of a minimally important change in VAS pain was a 30% or 20-point decrease from baseline. Secondary outcomes included the patient health questionnaire 9-item scale (PHQ-9, 0-27) for depression, the generalized anxiety disorder 7-item scale (GAD-7, 0-21) for anxiety, the work productivity and activity impairment (WPAI) scale, the knee injury and osteoarthritis outcome score—pain subscale (KOOS—pain, 100-0) for knee pathway participants, the Modified von Korff scale (MvK, 0-100) for back pathway participants [20,21], and surgery likelihood ("What do you think are the chances you'll have [back/knee] surgery in the next year, in %?", 0-100%). These secondary outcomes were collected at baseline, 6-weeks, and 12-weeks. Other baseline measurements obtained at week 0 consisted of participants' age, gender, and BMI. Participants' engagement with the DCP was measured by recording the number of ET sessions completed, the number of coaching interactions, and the number of education papers read. Each coaching interaction was further categorized as participant-to-coach or coach-to-participant; phone calls with a coach were not recorded as an interaction. Program satisfaction was asked at week 12 ("On a scale of 0-10, how likely is it that you would recommend the Hinge Health program to a friend or colleague?", 0-10).

Statistical Analysis

The distribution of gender and BMI in the knee and back pathways were compared using 2-sided Fisher's exact test and Mann-Whitney test, respectively. The association of baseline variables with program completion status was modeled using a logistic regression model and Wald's confidence intervals for the odds ratios (ORs) estimated. Exploratory analyses visualized the relationship between overall pain reduction over the course of the DCP and the total number of ET sessions (grouped in equisized bins assuming an average of 35 ET sessions for program completers). VAS pain trends were modeled using piecewise linear regression splines. Intersubject variability in the rate of change was modeled through random effects and used a first-order autoregression correlation structure to model within-subject correlation in residuals. Optimal knot locations for the spline were determined by a cross-validation procedure that evaluated model fit on a grid of knot locations. The fixed effects were estimated using a linear mixed-effects model (Multimedia Appendix 3). Significance (*P* value) evaluation was based on Wald *t* values with a Satterthwaite correction. For pain-response subgroup analysis, a Gaussian mixture model was fitted to the estimated spline coefficients to discover clusters corresponding to subgroups within the cohort, each with a distinct pain reduction trend. Adjusted ORs were computed to understand the association between participants' characteristics and the representative pain reduction trends for each subgroup. Finally, classification algorithms were trained to distinguish the 3 response groups based on the participants' demographic and baseline measurements alone, and performance was evaluated using 5-fold cross-validation. All analyses were performed using R statistical computing software.

Results

Participant Demographics and Digital Care Program Completion

Of the 10,264 DCP participants, 6468 self-reported back pain and were enrolled in the back-pain pathway and 3796 self-reported knee pain and were enrolled in the knee-pain pathway. The average age was 43.6 years, and the average BMI was 30.25. The proportion of female participants in the DCP was 50.00% (5132/10,264). Compared with the back-pain pathway, BMI was 1.3 kg/m^2 higher (P<.001) and the proportion of female participants was 3.9% higher (P<.001) in the knee-pain pathway. The difference in mean age between pathways was not significant (Table 1).

In all, 73.04% (7497/10,264) of the participants completed the DCP (referred to as *completers*), defined as completing at least one exercise session or reading 1 educational paper in weeks 9-12. Older users were more likely to complete the DCP (OR 1.037, 95% CI 1.03-1.04), whereas those with a higher BMI were less likely to complete the DCP (OR 0.973, 95% CI 0.97-0.98). No other baseline measures were significantly associated with completion (Multimedia Appendix 3). On average, completers engaged in 10.45 weeks with 35.02 ET

sessions and 19.39 education sessions. Table 2 summarizes the engagement by pathway for all participants and completers. No injuries or other adverse effects of DCP engagement other than temporary discomfort were reported.

Table 2. Mean engagement and SD for the full cohort and for completers by pathway (N=10,264).

Variables	All			Completers		
	Overall	Back pain	Knee pain	Overall	Back pain	Knee pain
Number of participants, n (%)	10,264 (100.00)	6468 (63.02)	3796 (36.98)	7497 (73.04)	4676 (72.29)	2821 (74.32)
Weeks engaged (ET[a] session or education session), mean (SD)	8.46 (3.9)	8.36 (3.92)	8.63 (3.86)	10.45 (2.15)	10.39 (2.17)	10.54 (2.1)
ET sessions per week, mean (SD)	2.93 (1.47)	2.85 (1.46)	3.05(1.47)	3.26 (1.39)	3.18 (1.41)	3.4 (1.34)
Total ET sessions, mean (SD)	27.43 (20.56)	26.48 (20.45)	29.04 (20.65)	35.02 (18.68)	34.04 (18.86)	36.65 (18.25)
Education sessions per week, mean (SD)	2.24 (1.55)	2.2 (1.55)	2.31 (1.56)	2.44 (1.28)	2.4 (1.27)	2.5 (1.3)
Total Education session, mean (SD)	15.33 (13.27)	14.81 (13.00)	16.24 (13.67)	19.39 (12.92)	18.84 (12.71)	20.29 (13.20)
Coach interactions per week, mean (SD)	7.03 (3.21)	6.99 (3.09)	7.09 (3.39)	7.23 (3.25)	7.21 (3.15)	7.27 (3.4)
Total coach interactions, mean (SD)	84.08 (43.3)	83.55 (42.02)	84.97 (45.36)	91.47 (43.42)	91.03 (42.33)	92.19 (45.16)

[a]ET: exercise therapy.

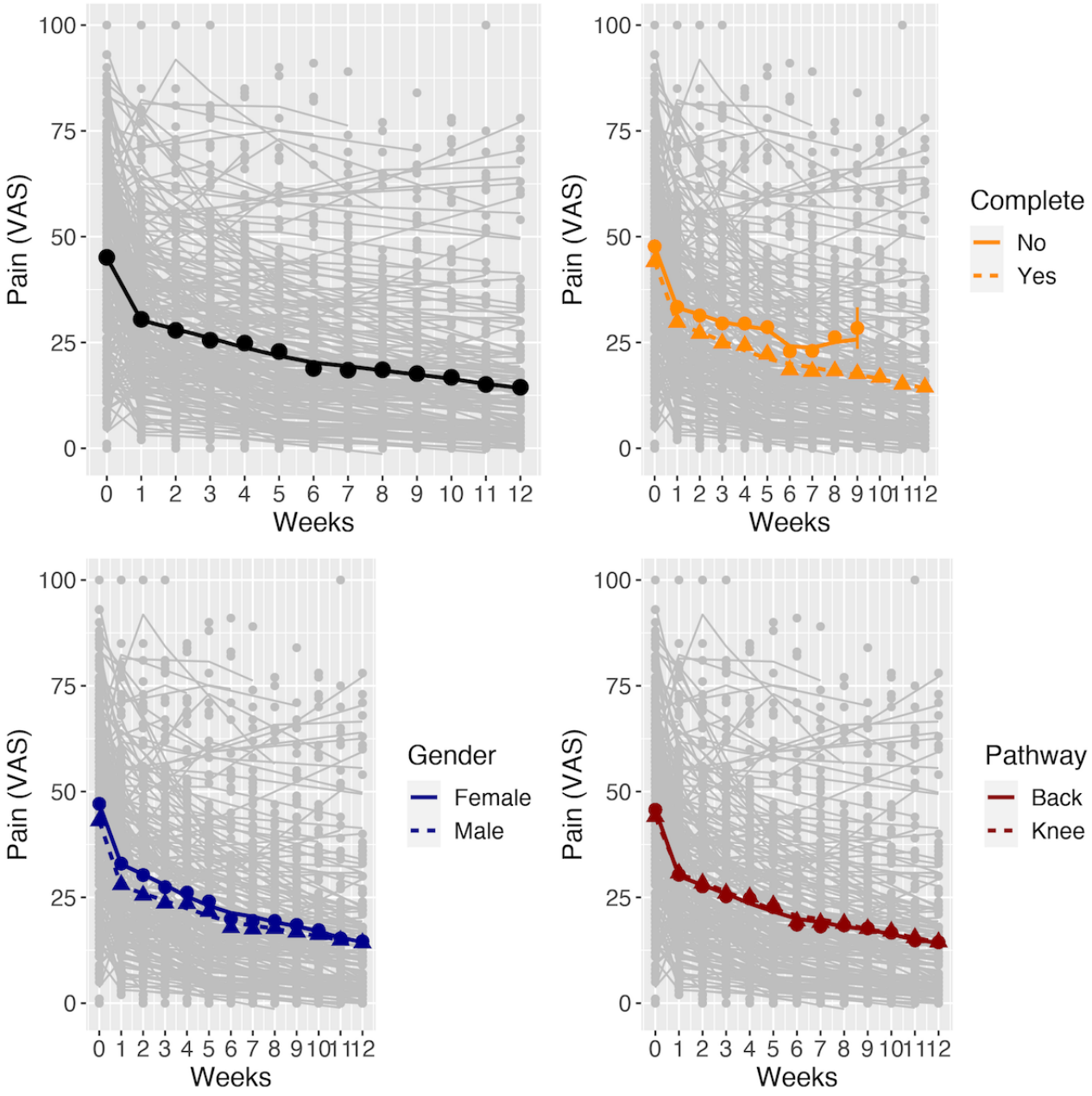
Longitudinal Changes in Pain

On the basis of a linear mixed effects model, the estimated mean reduction in pain by week 12 was 68.45% (30.89 points). Participants' pain scores changed nonlinearly over time (Figure 1). The mean change in pain scores per week (adjusted for sex, pathway, baseline age, BMI, anxiety, and depression scores) was 15.96 points for week 1 (P<.001) and 1.11 points per week for weeks 6-12 (P<.001) but was not significant for weeks 2-5. The conditional and marginal R-squared statistics [22] for our model were 0.94 and 0.54, respectively.

Minimally important change from baseline pain (defined as either a VAS pain reduction of 20 points or 30% with respect to baseline) was achieved by 78.60% (5893/7497) of completers and 69.60% (7144/10,264) of all participants.

Completers demonstrated greater pain reduction than noncompleters (Figure 1, top right) with an increased mean reduction rate of 0.48 points per week (SE 0.14) in weeks 2-5. Final pain reduction was nearly identical for both male and female genders (Figure 1, bottom left). However, there was a significantly higher mean reduction rate for male participants in the first week (mean difference=0.89 points per week, SE 0.46), and lower mean reduction rates in weeks 2-5 (difference=0.47 points per week, SE 0.09) and weeks 6-12 (difference=0.22 points per week, SE 0.05). Compared with the knee pathway, the back pathway was associated with a higher mean pain reduction rate (difference=3.1 points per week, SE 0.48) in the first week, but the pathway was not a significant variable in later weeks (Figure 1, bottom right).

Figure 1. Longitudinal changes in pain. The panels show the average pain scores computed for the entire study cohort (circles) and the fitted means (lines) computed for weeks 0-12 of the study. Top left shows the overall fitted mean. The plots on the top right, bottom left and bottom right show the means for subjects grouped by completion status, gender, and pathway, respectively. Weekly recorded pain and fitted curves for a random sample of subjects are plotted in gray on each panel. Error bars indicate 1 SE of the mean. F: female; M: male; VAS: visual analog scale for pain.

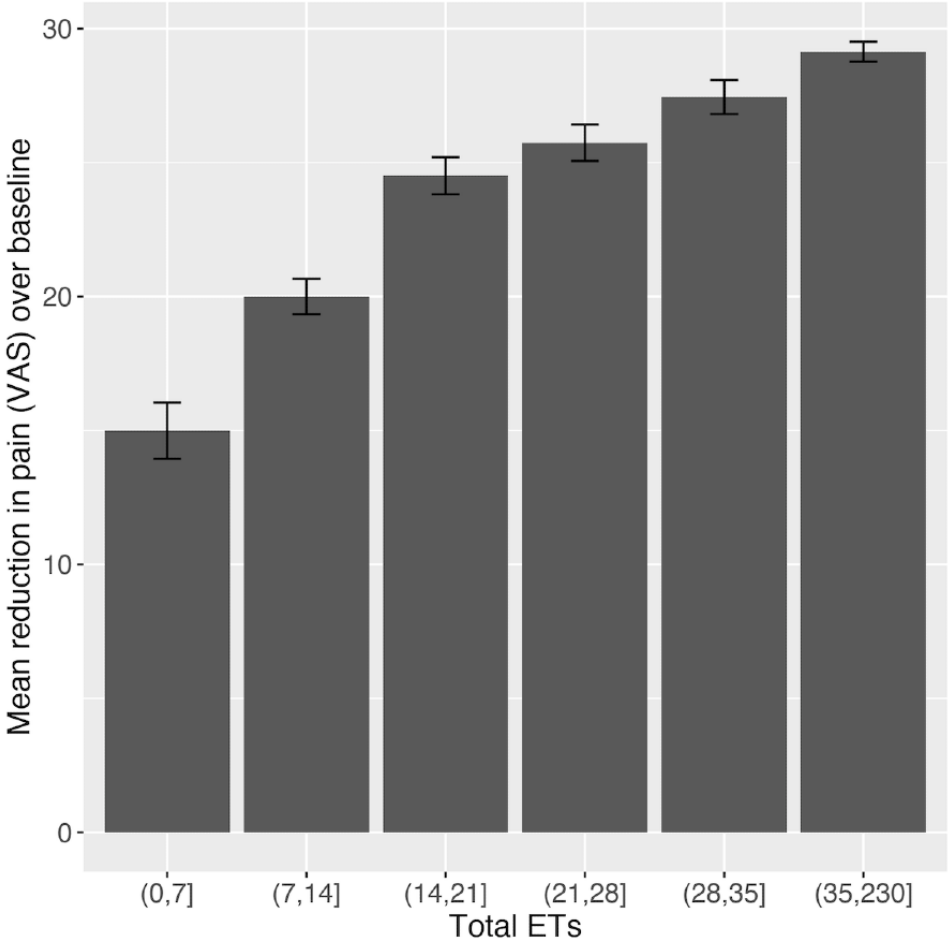


Effect From Engagement

Increasing levels of ET engagement in the DCP were associated with greater reductions in VAS pain score (P<.001; Figure 2). Notably, the relationship between the change in pain score and the number of ET sessions was nonlinear, with initial ET sessions contributing a higher proportion of the mean reduction achieved. The rate of reduction (adjusted for gender, pathway, baseline age, BMI, anxiety, and depression scores) for the initial 10 ET sessions was 1.9 VAS points per session (SE 0.2; P<.001).

The number of weekly coach interactions was also associated with a reduction in pain with a mean reduction of 0.18 VAS points per interaction (SE 0.06; P=.003) for the first 30 interactions. The number of participant-to-coach interactions, specifically, was associated with a mean rate of reduction in pain of 0.30 VAS points per interaction (SE 0.1; P=.003) for the first 20 interactions. The number of coach-to-participant interactions was not significantly associated with pain reduction.

Figure 2. Association between pain reduction and ET sessions. Bar plots show the mean reduction in pain achieved over the DCP grouped by the total number of ET sessions. Error bars indicate 1 SE of the mean. DCP: digital care pathway; ET: exercise therapy; VAS: visual analog scale for pain.



Mental Health and Other Secondary Outcome Measures

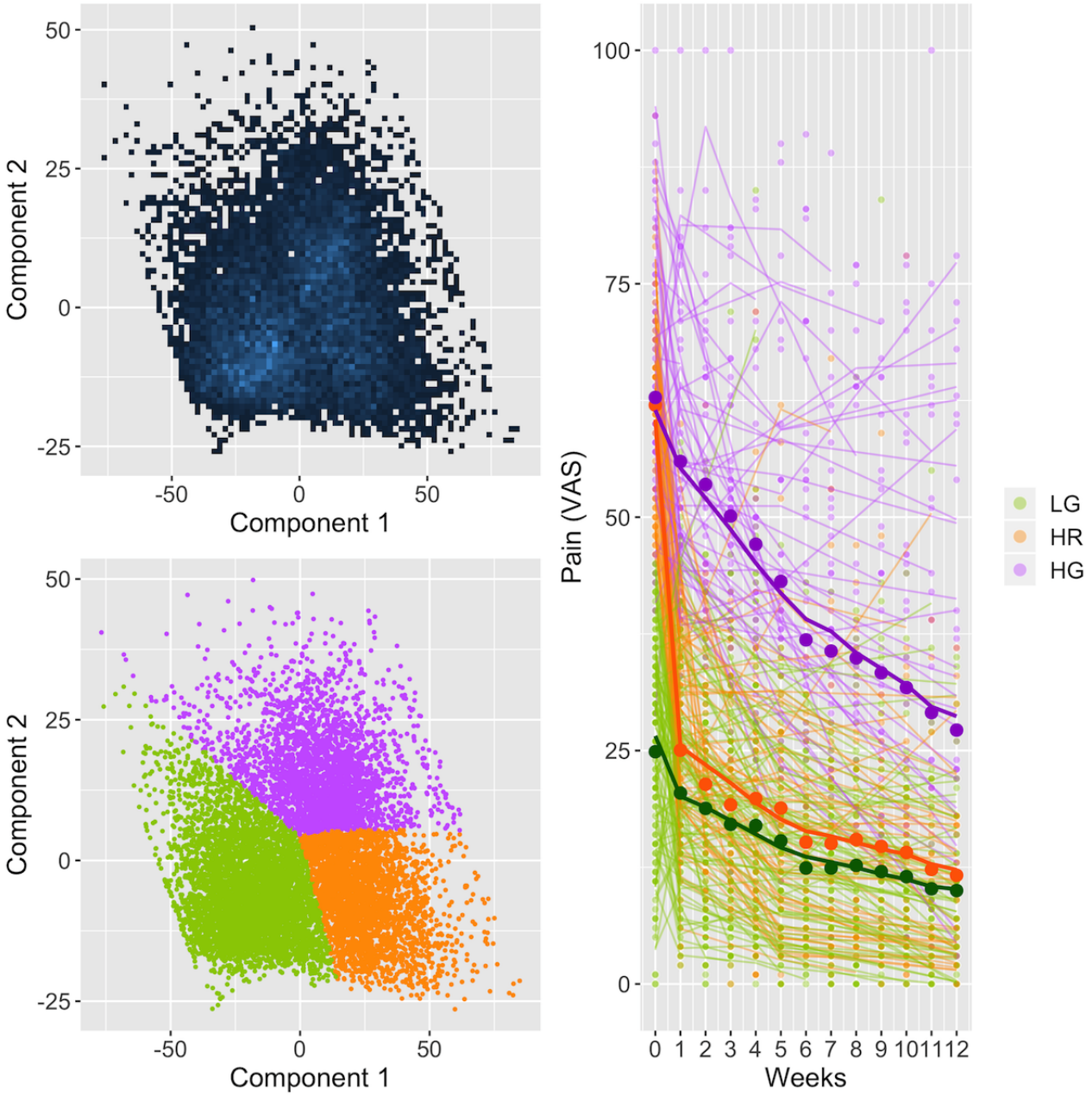
For participants categorized as having depressive symptoms (PHQ-9≥5) at baseline, the mean baseline PHQ-9 score was 12.01 and decreased by 57.9% to 5.05 at week 11 (P<.001). Differences between the pathways were not significant. The percentage of patients with depressive symptoms at baseline and at the end of the study was 21.1% and 11.4%, respectively. For participants categorized as having anxiety symptoms (GAD-7≥5) at baseline, the mean baseline GAD-7 score was 11.49 and decreased by 58.3% to 4.78 at week 11 (P<.001). The back pathway participants had a 0.46 point (P<.001) greater mean GAD-7 reduction than those in the knee pathway. The percentage of patients with anxiety symptoms at baseline and the end of the study was 28.3% and 14.2%, respectively (PHQ-9 and GAD-7 values at week 6 were carried forward to impute missing values at week 12).

With respect to baseline, the mean surgery likelihood score decreased by 67.4% (8.15 points, P<.001) overall, and by 66.8% and 68.2% for knee and back pathway participants, respectively. The mean KOOS—pain decreased by 33.9% (5.19 points, P<.001) in knee pathway participants and the mean MvK decreased by 51.4% (8.20 points, P<.001) in the back pathway participants. The within-participant correlation coefficients for KOOS—pain and MvK scores (with VAS pain) were 0.59 (95% CI 0.58-0.61) and 0.80 (95% CI 0.79, 0.81), respectively, indicating strong correlations between the primary and secondary pain variables. The mean WPAI score decreased by 63.94% from baseline (20.29 points, P<.001). The DCP final satisfaction score was 8.97/10 with a net promoter score of +64/100.

Distinct Pain Response Groups

Intersubject variation in pain reduction trends motivated a subgroup analysis of pain response, and 3 distinct response groups emerged (Figure 3). Participants with high pain at baseline and gradual improvement were designated as *high gradual* (HG). Participants with high baseline pain but a rapid decline were labeled *high rapid* (HR), and those with low baseline pain and gradual response were labeled *low gradual* (LG). All LG participants had baseline pain below 50. HR participants had the highest mean pain reduction over the duration of the DCP (48.8 points, 80.0%), followed by the HG (33.3 points, 54.1%) and LG group (15.3 points, 64.0%).

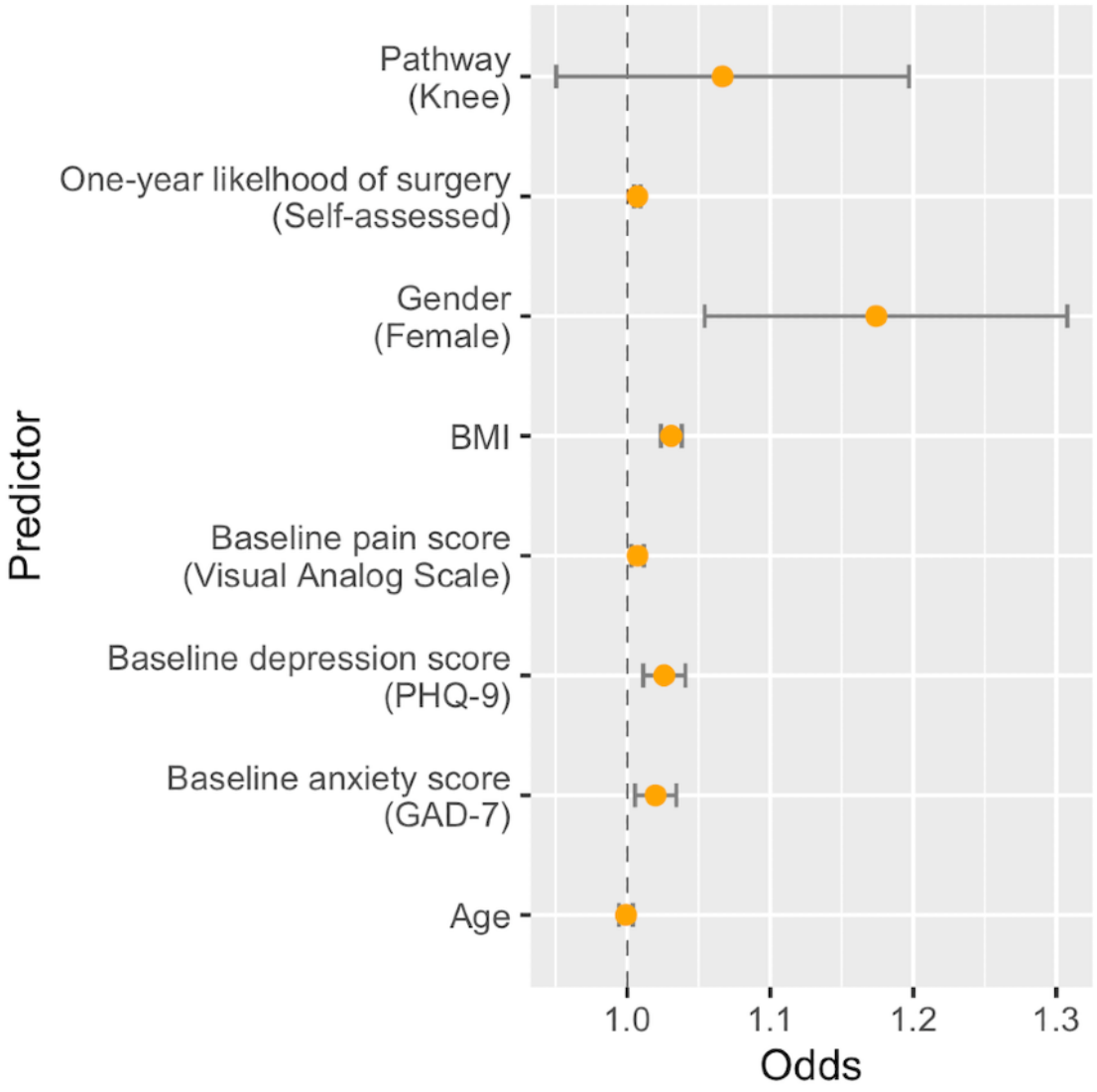
Figure 3. Pain response subgroups. Pain reduction trend clusters obtained by fitting a 3-component GMM identified 3 subgroups (HG, HR, and LG response). (Top left) 2D density plot of the first 2 principal components of the fitted splines shows each of the 3 subgroups. (Bottom left) Curves denoted by their respective principal components 1 and 2 are assigned to a cluster based on maximum posterior likelihood. (Right) Random sample of pain reduction trends colored by subgroup and the respective mean trends. 2D, 2 dimensional; GMM, Gaussian Mixture Model; HG, high-gradual; HR, high-rapid; LG: low-gradual.



Relative to the HR response, female participants had 17.3% (*P*=.002) higher odds of an HG response (Figure 4). The odds of an HG response also increased by 3.1% (*P*<.001) per unit increase in BMI and increased by 2.2% (*P*=.001) and 2.1% (*P*=.002) per unit increase in PHQ-9 and GAD-7, respectively.

Classification of response groups based on baseline attributes achieved a mean accuracy of 76% (SE 0.3%) using a random forest algorithm, evaluated using 5-fold cross-validation. The classifier had a mean area under the precision-recall curve of 68.92% (SE 2.04%). Nearly equal numbers of participants belonged to the HR and HG response groups. Subgroup analysis details are provided in Multimedia Appendix 3.

Figure 4. Association of baseline variables with a high-gradual or high-rapid pain reduction trend. For each baseline variable, the plotted values indicate the odds ratios for a gradual response (with reference to a high-rapid response) for a unit increase in the corresponding predictor. The error bars denote the 95% profile-likelihood CIs.



Discussion

Principal Findings

This study demonstrated the positive effect of a 12-week DCP on chronic musculoskeletal pain outcomes in a large sample of real-world patients. Specifically, participants experienced a 68.5% average improvement in VAS pain between baseline and 12 weeks, and 78.60% (5893/7497) of program completers (7144/10,264, 69.60% of all participants) achieved clinically meaningful improvement. Completion was high, with 73.04% (7497/10,264) of participants reaching the final month, and completers engaged in a mean of 35.0 ET sessions, 19.4 education sessions, and 91.5 coach interactions during the DCP. It is well known that unless a digital health intervention fits into users' daily lives, only a small proportion of all participants who sign up actually complete the program [23,24]. The exceptional completion rate of our study may be due to the multipronged strategy of our DCP that uses both a digital and a human interface to engage with participants. Furthermore, both the number of ET sessions and participant-to-coach interactions were positively associated with improvement in pain, supporting that the level of participant engagement influenced outcomes. These results support the effectiveness of a DCP for musculoskeletal pain in the real-world setting, and the large sample size supports the prospect for scalability to serve a large number of chronic low back and knee pain sufferers worldwide.

Comparison With Literature

The observed 68.5% average improvement in VAS pain in this DCP outperforms the pain reduction effect sizes observed in a variety of conservative care interventions with similar timeframes. For pain associated with knee osteoarthritis, comparable conservative care interventional studies demonstrate an average improvement in VAS pain of 19%-48% [25-28]. For low back pain, comparable studies demonstrate average improvements in VAS pain of 29%-53% [29-34]. Similarly, a systematic review of randomized clinical trials for low back pain showed a within-group standardized mean difference of 1.07 (95% CI 0.87-1.27) for pain reduction at 13 weeks [35], whereas a standardized mean difference of 1.37 (95% CI

1.33-1.40) for pain reduction at 12 weeks was observed in this study. Not only does this study demonstrate greater improvement in pain for both knee and low back pathways but it also has a much larger sample size than previous studies, which typically did not exceed 100 subjects. Furthermore, this study found a strong correlation between changes in VAS pain and secondary pain measures (KOOS—pain for knee and MvK for back), further supporting the validity of the VAS pain measurements. Finally, compared with other studies utilizing therapeutic exercise for chronic pain, this study demonstrated a similar lack of adverse events. This is likely attributable to the benefits and safety of light intensity stretching and strengthening exercises, and in this study may also be due to the exercise guidance provided by the wearable sensors.

Patient Engagement

Notably, most previous studies have occurred in traditional clinical settings, where multiple barriers prevent both patients and clinicians from engaging in conservative care [36]. For example, adherence of chronic low back pain patients to home exercises prescribed from traditional physical therapy ranges from 30% to 50% and remains a significant challenge for administering effective care [37,38]. A primary benefit of a digital care approach for chronic musculoskeletal pain is the ability to engage patients with their treatment and self-management. Smartphone apps can cost-effectively deliver education and encourage healthy behaviors, whereas sensors can provide exercise guidance and track engagement [39]. The DCP in this study engaged 73.04% (7497/10,264) of users to completion, with completers engaging in 10.5 of the 12 weeks, including 3.3 ET sessions, 2.4 education papers, and 7.2 coach interactions per week (mean ET sessions 2.9, mean education sessions 2.2, and 7.0 coach interactions per week among all participants). Notably, this study demonstrated an association between pain improvement and both the number of ET sessions and the number of coach interactions, suggesting that the level of participant engagement impacted the results. Specifically, the first 10 ET sessions and the first 30 coach interactions were the most influential in pain improvement. Of note, a recent study evaluating a DCP in a similar population showed lower engagement and no relationship between exercise and pain reduction, suggesting that specific program implementation details (ie, sensor-guided exercises and health coaching) may have a large effect on outcomes [40].

Mental Health Outcomes

Depression and anxiety are known to often occur in patients with chronic musculoskeletal pain [41], so the effects of this DCP on symptoms of depression and anxiety were also assessed. Behavioral health coaching and education on cognitive behavioral therapy concepts were key elements of the DCP's multimodal digital care approach. A large body of research confirms the effect of psychological factors, such as depression and anxiety, on chronic pain [42,43]. In particular, an association between chronic low back pain and psychological factors has been shown, and related therapeutic approaches, including cognitive behavioral therapy and mindfulness-based stress reduction, have demonstrated effectiveness for back pain reduction [44]. This study showed that outcomes for participants with symptoms of depression and anxiety decreased on average by 57.9% and 58.3%, respectively, over the course of the DCP. This suggests a strong relationship between mental health and pain improvement; however, a causal relationship between these entities cannot be determined. Notably, mental health improvements were very similar across knee and back pathways, whereas a small difference (0.46 points) in GAD-7 outcomes was noted. This is unlikely to be clinically meaningful. Future work will further explore the effect of coaching and other behavioral health support on pain and functional outcomes.

Predicting Pain Response

In addition to clinical effectiveness, another potential benefit of a DCP is the insight gained from longitudinal tracking of outcome data in large populations. The large sample size in this study, combined with data collection at regular and relatively frequent time intervals, enabled the discovery of distinct clusters of pain response trajectories over time. Participants were classified as gradual versus rapid pain responders, and patient-specific features that influenced the likelihood of pain response category were identified. By clustering distinct trends in pain response over time for each subject, we specifically uncovered 3 distinct pain response subgroups. Two groups had high baseline pain but differed in the rate of recovery (rapid vs gradual), whereas the third group had low baseline pain with gradual recovery. Notably, we were able to forecast with 76% accuracy which of these pain response groups a user would fall into based on their baseline information. Looking specifically at the 2 groups with high baseline pain, the rapid response was more likely to occur in male participants, those with lower BMI, or those with lower depression or anxiety scores. These pain response groups enable a better understanding of temporal changes in pain during the rehabilitation process and may ultimately help to identify pain recovery mechanisms. Furthermore, continued research into response patterns may ultimately allow for a more personalized approach to care, including more accurate prognosis and additional treatment options for patients likely to have a more gradual recovery.

Strengths and Limitations

This study has several limitations, including the lack of a control group and the lack of physical function outcomes. Notably, previous randomized trials of this DCP on smaller populations (N<200) demonstrated positive effects on pain and functional outcomes (Oswestry disability index, KOOS—physical function short form) compared with control groups [17,18]. This study assessed outcomes in a sample of more than 10,000 users and demonstrated similar effectiveness. Another limitation of this study is the lack of long-term outcomes, and future studies should assess if participants are able to sustain healthy behaviors and self-management promoted in the DCP. Some potentially important demographic variables (ie, education, ethnicity, income, and smoking status) and medical history variables (ie, diabetes, hypertension, and mental health) were not obtained. Finally, this study was conducted through employers, which limits the applicability to clinical settings with higher proportions of uninsured, elderly, or work-disabled patients. However, this study was conducted with employees from 30 different companies across the United States and included a

wide diversity of job types (eg, truck drivers, manual laborers, office workers), suggesting that the findings are applicable to a broad population. In addition, older patients were more likely to complete the program than younger ones, emphasizing that digital health tools are not only useful to the younger population.

The strengths of this study include the large sample size in the real-world setting, which demonstrated scalability and enabled the discovery of unique features, such as distinct pain response clusters in longitudinal real-world data. In addition, this study had similar age and sex distributions for knee and back pain participants, enabling direct comparison of the separate knee and back pathways. The average pain response for these separate pathways was quite similar (Figure 1), which is notable given the assumed underlying pathological differences between knee and back pain, but supports recent work urging practitioners to move beyond separating body regions when managing chronic musculoskeletal pain [45]. Finally, this study demonstrates significant improvements in self-reported workplace productivity (WPAI, 61.5% improvement) and surgery likelihood (67.4% reduction), suggesting that a DCP may have considerable economic benefits.

Future Directions

DCPs may ultimately be used to complement clinical musculoskeletal practice, and further research is warranted on their use by patients and providers. This study supports the efficacy and scalability of a DCP for facilitating safe conservative care and promoting healthy behavior change. However, critical reviews have identified a lack of external and long-term validation of digital health tools [46]. Many previous studies on digital interventions for chronic low back pain have presented unconvincing results [47]. Given that digital health tools are typically developed in the private sector, and good clinical research can be time-consuming and challenging, we see a need for collaborative efforts between industry and academic medicine to optimize digital health technologies for effective conservative care implementation, adoption, and access in the broad, real-world population with musculoskeletal pain.

Conclusions

This study supports the efficacy and scalability of a DCP for chronic low back and knee pain in a large, real-world population. Participants demonstrated very high completion and engagement rates, and a significant positive relationship between engagement and pain reduction was identified. This is the first longitudinal digital health study to analyze musculoskeletal health outcomes in a sample of this magnitude, and it supports the prospect for DCP scalability to serve the overwhelming number of chronic back and knee pain sufferers worldwide. Furthermore, the large sample size enabled the prediction of rapid versus gradual pain response from baseline information, which may prove beneficial for prognosis and tailoring future interventions.

Acknowledgments

The sponsors of the study, Hinge Health, Inc, facilitated data collection. The sponsor and all authors participated in the decision to submit the paper for publication.

Authors' Contributions

JB, VA, and JK contributed to the study design, analysis plan, interpretation, and manuscript drafting. JB and VA contributed equally with JB leading analysis strategy and interpretation for clinical impact, VA conducting statistical analysis, and each drafting half of the manuscript. MS, PZ, DK, and MF contributed to manuscript drafting and editing. All the authors had access to all the raw data. JB and VA led all data analysis, data interpretation, and writing of the report.

Conflicts of Interest

JB and VA were academic collaborators in this study and were paid as consultants for their expertise in conducting this study. JB and VA had unrestricted access to the data and led the analysis and interpretation of the results. Subsequent to paper completion, author VA became employed at Hinge Health, Inc, and received salary and equity compensation. Author JK is employed at Hinge Health, Inc, and receives salary and equity compensation. MF and DK are clinical advisors to Hinge Health, Inc, and receive equity compensation. Authors PZ and MS have no relevant declarations.

Multimedia Appendix 1
Hinge Health Digital Care Pathway.
[PDF File (Adobe PDF File), 41435 KB - jmir_v22i5e18250_app1.pdf]

Multimedia Appendix 2
Hinge Health Digital Back Pathway - woodpecker exercise.
[MP4 File (MP4 Video), 24823 KB - jmir_v22i5e18250_app2.mp4]

Multimedia Appendix 3
Supplementary tables and figures.
[DOCX File , 260 KB - jmir_v22i5e18250_app3.docx]

References

1. Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the global burden of disease study 2010. Lancet 2012 Dec 15;380(9859):2197-2223. [doi: 10.1016/S0140-6736(12)61689-4] [Medline: 23245608]
2. Kotlarz H, Gunnarsson CL, Fang H, Rizzo JA. Insurer and out-of-pocket costs of osteoarthritis in the US: evidence from national survey data. Arthritis Rheum 2009 Dec;60(12):3546-3553 [FREE Full text] [doi: 10.1002/art.24984] [Medline: 19950287]
3. United States Bone and Joint Decade. Burden of Musculoskeletal Diseases in the United States: Prevalence, Societal and Economic Cost. Second Edition. Rosemont, IL: American Academy of Orthopaedic Surgeons; 2010.
4. Ringwalt C, Gugelmann H, Garrettson M, Dasgupta N, Chung AE, Proescholdbell SK, et al. Differential prescribing of opioid analgesics according to physician specialty for medicaid patients with chronic noncancer pain diagnoses. Pain Res Manag 2014;19(4):179-185 [FREE Full text] [doi: 10.1155/2014/857952] [Medline: 24809067]
5. Wright BJ, Galtieri NJ, Fell M. Non-adherence to prescribed home rehabilitation exercises for musculoskeletal injuries: the role of the patient-practitioner relationship. J Rehabil Med 2014 Feb;46(2):153-158 [FREE Full text] [doi: 10.2340/16501977-1241] [Medline: 24322559]
6. Reuben DB, Alvanzo AA, Ashikaga T, Bogat GA, Callahan CM, Ruffing V, et al. National institutes of health pathways to prevention workshop: the role of opioids in the treatment of chronic pain. Ann Intern Med 2015 Feb 17;162(4):295-300. [doi: 10.7326/M14-2775] [Medline: 25581341]
7. Qaseem A, Wilt TJ, McLean RM, Forciea MA, Clinical Guidelines Committee of the American College of Physicians. Noninvasive treatments for acute, subacute, and chronic low back pain: a clinical practice guideline from the American college of physicians. Ann Intern Med 2017 Apr 4;166(7):514-530. [doi: 10.7326/M16-2367] [Medline: 28192789]
8. Moseley JB, O'Malley K, Petersen NJ, Menke TJ, Brody BA, Kuykendall DH, et al. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. N Engl J Med 2002 Jul 11;347(2):81-88. [doi: 10.1056/NEJMoa013259] [Medline: 12110735]
9. Delitto A, Piva SR, Moore CG, Fritz JM, Wisniewski SR, Josbeno DA, et al. Surgery versus nonsurgical treatment of lumbar spinal stenosis: a randomized trial. Ann Intern Med 2015 Apr 7;162(7):465-473 [FREE Full text] [doi: 10.7326/M14-1420] [Medline: 25844995]
10. Slattery BW, Haugh S, O'Connor L, Francis K, Dwyer CP, O'Higgins S, et al. An evaluation of the effectiveness of the modalities used to deliver electronic health interventions for chronic pain: systematic review with network meta-analysis. J Med Internet Res 2019 Jul 17;21(7):e11086 [FREE Full text] [doi: 10.2196/11086] [Medline: 31317869]
11. Dorflinger L, Kerns RD, Auerbach SM. Providers' roles in enhancing patients' adherence to pain self management. Transl Behav Med 2013 Mar;3(1):39-46 [FREE Full text] [doi: 10.1007/s13142-012-0158-z] [Medline: 24073159]
12. Auerbach AD. Evaluating digital health tools-prospective, experimental, and real world. JAMA Intern Med 2019 Jun 1;179(6):840-841. [doi: 10.1001/jamainternmed.2018.7229] [Medline: 30801610]
13. Berman MA, Guthrie NL, Edwards KL, Appelbaum KJ, Njike VY, Eisenberg DM, et al. Change in glycemic control with use of a digital therapeutic in adults with type 2 diabetes: cohort study. JMIR Diabetes 2018 Feb 14;3(1):e4 [FREE Full text] [doi: 10.2196/diabetes.9591] [Medline: 30291074]
14. Milani RV, Lavie CJ, Bober RM, Milani AR, Ventura HO. Improving hypertension control and patient engagement using digital tools. Am J Med 2017 Jan;130(1):14-20. [doi: 10.1016/j.amjmed.2016.07.029] [Medline: 27591179]
15. Espie CA, Pawlecki B, Waterfield D, Fitton K, Radocchia M, Luik AI. Insomnia symptoms and their association with workplace productivity: cross-sectional and pre-post intervention analyses from a large multinational manufacturing company. Sleep Health 2018 Jun;4(3):307-312. [doi: 10.1016/j.sleh.2018.03.003] [Medline: 29776626]
16. Cronström A, Dahlberg LE, Nero H, Hammarlund CS. 'I was considering surgery because I believed that was how it was treated': a qualitative study on willingness for joint surgery after completion of a digital management program for osteoarthritis. Osteoarthritis Cartilage 2019 Jul;27(7):1026-1032 [FREE Full text] [doi: 10.1016/j.joca.2019.04.004] [Medline: 31002940]
17. Mecklenburg G, Smittenaar P, Erhart-Hledik JC, Perez DA, Hunter S. Effects of a 12-week digital care program for chronic knee pain on pain, mobility, and surgery risk: randomized controlled trial. J Med Internet Res 2018 Apr 25;20(4):e156 [FREE Full text] [doi: 10.2196/jmir.9667] [Medline: 29695370]
18. Shebib R, Bailey JF, Smittenaar P, Perez DA, Mecklenburg G, Hunter S. Randomized controlled trial of a 12-week digital care program in improving low back pain. NPJ Digit Med 2019;2:1 [FREE Full text] [doi: 10.1038/s41746-018-0076-7] [Medline: 31304351]
19. Tomkins-Lane CC, Lafave LM, Parnell JA, Rempel J, Moriartey S, Andreas Y, et al. The spinal stenosis pedometer and nutrition lifestyle intervention (SSPANLI): development and pilot. Spine J 2015 Apr 1;15(4):577-586. [doi: 10.1016/j.spinee.2014.10.015] [Medline: 25452012]
20. Underwood MR, Barnett AG, Vickers MR. Evaluation of two time-specific back pain outcome measures. Spine (Phila Pa 1976) 1999 Jun 1;24(11):1104-1112. [doi: 10.1097/00007632-199906010-00010] [Medline: 10361660]
21. von Korff M, Ormel J, Keefe FJ, Dworkin SF. Grading the severity of chronic pain. Pain 1992 Aug;50(2):133-149. [doi: 10.1016/0304-3959(92)90154-4] [Medline: 1408309]

22. Nakagawa S, Schielzeth H. A general and simple method for obtaining R2 from generalized linear mixed ‐ effects models. Methods Ecol Evol 2013;4(2):133-142 [FREE Full text] [doi: 10.1111/j.2041-210x.2012.00261.x]
23. 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]
24. Anhøj J, Jensen AH. Using the internet for life style changes in diet and physical activity: a feasibility study. J Med Internet Res 2004 Sep 8;6(3):e28 [FREE Full text] [doi: 10.2196/jmir.6.3.e28] [Medline: 15471754]
25. Fransen M, McConnell S, Harmer AR, van der Esch M, Simic M, Bennell KL. Exercise for osteoarthritis of the knee: a Cochrane systematic review. Br J Sports Med 2015 Dec;49(24):1554-1557. [doi: 10.1136/bjsports-2015-095424] [Medline: 26405113]
26. Bennell KL, Hinman RS, Metcalf BR, Buchbinder R, McConnell J, McColl G, et al. Efficacy of physiotherapy management of knee joint osteoarthritis: a randomised, double blind, placebo controlled trial. Ann Rheum Dis 2005 Jun;64(6):906-912 [FREE Full text] [doi: 10.1136/ard.2004.026526] [Medline: 15897310]
27. Doi T, Akai M, Fujino K, Iwaya T, Kurosawa H, Hayashi K, et al. Effect of home exercise of quadriceps on knee osteoarthritis compared with nonsteroidal antiinflammatory drugs: a randomized controlled trial. Am J Phys Med Rehabil 2008 Apr;87(4):258-269. [doi: 10.1097/PHM.0b013e318168c02d] [Medline: 18356618]
28. Lund H, Weile U, Christensen R, Rostock B, Downey A, Bartels EM, et al. A randomized controlled trial of aquatic and land-based exercise in patients with knee osteoarthritis. J Rehabil Med 2008 Feb;40(2):137-144 [FREE Full text] [doi: 10.2340/16501977-0134] [Medline: 18509579]
29. Yelland MJ, Glasziou PP, Bogduk N, Schluter PJ, McKernon M. Prolotherapy injections, saline injections, and exercises for chronic low-back pain: a randomized trial. Spine (Phila Pa 1976) 2004 Jan 1;29(1):9-16; discussion 16. [doi: 10.1097/01.BRS.0000105529.07222.5B] [Medline: 14699269]
30. Doğan SK, Tur BS, Kurtaiş Y, Atay MB. Comparison of three different approaches in the treatment of chronic low back pain. Clin Rheumatol 2008 Jul;27(7):873-881. [doi: 10.1007/s10067-007-0815-7] [Medline: 18188660]
31. Saper RB, Lemaster C, Delitto A, Sherman KJ, Herman PM, Sadikova E, et al. Yoga, physical therapy, or education for chronic low back pain: a randomized noninferiority trial. Ann Intern Med 2017 Jul 18;167(2):85-94 [FREE Full text] [doi: 10.7326/M16-2579] [Medline: 28631003]
32. Goldby LJ, Moore AP, Doust J, Trew ME. A randomized controlled trial investigating the efficiency of musculoskeletal physiotherapy on chronic low back disorder. Spine (Phila Pa 1976) 2006 May 1;31(10):1083-1093. [doi: 10.1097/01.brs.0000216464.37504.64] [Medline: 16648741]
33. Critchley DJ, Ratcliffe J, Noonan S, Jones RH, Hurley MV. Effectiveness and cost-effectiveness of three types of physiotherapy used to reduce chronic low back pain disability: a pragmatic randomized trial with economic evaluation. Spine (Phila Pa 1976) 2007 Jun 15;32(14):1474-1481. [doi: 10.1097/BRS.0b013e318067dc26] [Medline: 17572614]
34. Williams KA, Petronis J, Smith D, Goodrich D, Wu J, Ravi N, et al. Effect of Iyengar yoga therapy for chronic low back pain. Pain 2005 May;115(1-2):107-117. [doi: 10.1016/j.pain.2005.02.016] [Medline: 15836974]
35. Artus M, van der Windt DA, Jordan KP, Hay EM. Low back pain symptoms show a similar pattern of improvement following a wide range of primary care treatments: a systematic review of randomized clinical trials. Rheumatology (Oxford) 2010 Dec;49(12):2346-2356. [doi: 10.1093/rheumatology/keq245] [Medline: 20713495]
36. Finnerup NB. Nonnarcotic methods of pain management. N Engl J Med 2019 Jun 20;380(25):2440-2448. [doi: 10.1056/NEJMra1807061] [Medline: 31216399]
37. Friedrich M, Cermak T, Maderbacher P. The effect of brochure use versus therapist teaching on patients performing therapeutic exercise and on changes in impairment status. Phys Ther 1996 Oct;76(10):1082-1088. [doi: 10.1093/ptj/76.10.1082] [Medline: 8863761]
38. Medina-Mirapeix F, Escolar-Reina P, Gascón-Cánovas JJ, Montilla-Herrador J, Jimeno-Serrano FJ, Collins SM. Predictive factors of adherence to frequency and duration components in home exercise programs for neck and low back pain: an observational study. BMC Musculoskelet Disord 2009 Dec 9;10:155 [FREE Full text] [doi: 10.1186/1471-2474-10-155] [Medline: 19995464]
39. Gal R, May AM, van Overmeeren EJ, Simons M, Monninkhof EM. The effect of physical activity interventions comprising wearables and smartphone applications on physical activity: a systematic review and meta-analysis. Sports Med Open 2018 Sep 3;4(1):42 [FREE Full text] [doi: 10.1186/s40798-018-0157-9] [Medline: 30178072]
40. Toelle TR, Utpadel-Fischler DA, Haas K, Priebe JA. App-based multidisciplinary back pain treatment versus combined physiotherapy plus online education: a randomized controlled trial. NPJ Digit Med 2019;2:34 [FREE Full text] [doi: 10.1038/s41746-019-0109-x] [Medline: 31304380]
41. Pinheiro MB, Morosoli JJ, Colodro-Conde L, Ferreira PH, Ordoñana JR. Genetic and environmental influences to low back pain and symptoms of depression and anxiety: a population-based twin study. J Psychosom Res 2018 Feb;105:92-98. [doi: 10.1016/j.jpsychores.2017.12.007] [Medline: 29332639]
42. Bair MJ, Robinson RL, Katon W, Kroenke K. Depression and pain comorbidity: a literature review. Arch Intern Med 2003 Nov 10;163(20):2433-2445. [doi: 10.1001/archinte.163.20.2433] [Medline: 14609780]
43. Surah A, Baranidharan G, Morley S. Chronic pain and depression. Cont Educ Anaesth Crit Care Pain 2014 Apr;14(2):85-89. [doi: 10.1093/bjaceaccp/mkt046]

44. Cherkin DC, Herman PM. Cognitive and mind-body therapies for chronic low back pain and neck pain: effectiveness and value. JAMA Intern Med 2018 Apr 1;178(4):556-557. [doi: 10.1001/jamainternmed.2018.0113] [Medline: 29507946]
45. Caneiro JP, Roos EM, Barton CJ, O'Sullivan K, Kent P, Lin I, et al. It is time to move beyond 'body region silos' to manage musculoskeletal pain: five actions to change clinical practice. Br J Sports Med 2019 Oct 11:- epub ahead of print. [doi: 10.1136/bjsports-2018-100488] [Medline: 31604698]
46. Sundararaman LV, Edwards RR, Ross EL, Jamison RN. Integration of mobile health technology in the treatment of chronic pain: a critical review. Reg Anesth Pain Med 2017;42(4):488-498. [doi: 10.1097/AAP.0000000000000621] [Medline: 28570436]
47. Nicholl BI, Sandal LF, Stochkendahl MJ, McCallum M, Suresh N, Vasseljen O, et al. Digital support interventions for the self-management of low back pain: a systematic review. J Med Internet Res 2017 May 21;19(5):e179 [FREE Full text] [doi: 10.2196/jmir.7290] [Medline: 28550009]

Abbreviations

DCP: digital care program
ET: exercise therapy
GAD-7: generalized anxiety disorder 7-item
HG: high gradual
HR: high rapid
KOOS—pain: knee injury and osteoarthritis outcome score—pain
LG: low gradual
MvK: Modified von Korff scale
OR: odds ratio
PHQ-9: patient health questionnaire 9-item scale
VAS: visual analog scale
WPAI: work productivity and activity impairment

Edited by G Eysenbach; submitted 19.02.20; peer-reviewed by D Goyal, R Milani; comments to author 11.03.20; revised version received 21.03.20; accepted 21.03.20; published 11.05.20.

Please cite as:
Bailey JF, Agarwal V, Zheng P, Smuck M, Fredericson M, Kennedy DJ, Krauss J
Digital Care for Chronic Musculoskeletal Pain: 10,000 Participant Longitudinal Cohort Study
J Med Internet Res 2020;22(5):e18250
URL: https://www.jmir.org/2020/5/e18250
doi:10.2196/18250
PMID:32208358

Original Paper

Revisiting Effective Communication Between Patients and Physicians: Cross-Sectional Questionnaire Study Comparing Text-Based Electronic Versus Face-to-Face Communication

Tala Mirzaei[1], PhD; Nicole Kashian[2], PhD

[1]Department of Information Systems and Business Analytics, Florida International University, Miami, FL, United States

[2]Department of Communication, Florida International University, Miami, FL, United States

Corresponding Author:
Tala Mirzaei, PhD
Department of Information Systems and Business Analytics
Florida International University
11200 SW 8th Street
Miami, FL, 33199-0001
United States
Phone: 1 3053482830
Email: tmirzaei@fiu.edu

Abstract

Background: Research has shown that text-based communication via telemedicine will continue to be a mode of communication that patients and physicians use in the future. However, very few studies have examined patients' perspectives regarding the increased use of text-based communication versus face-to-face (FtF) communication.

Objective: This study aimed to understand and compare the potential differences in patients' perceptions of communication effectiveness with their physicians through different modes of communication.

Methods: We conducted a web-based survey of 345 patients to explore the impact of different channels on effective communication and perceived health behavior and outcomes. We tested the impact of patients' perceived communication and media effectiveness on their self-efficacy, communication satisfaction, and perceived health outcomes, separately for text-based information technology (IT)–mediated communication and FtF communication. Furthermore, we conducted a group comparison to identify significant differences across these 2 groups.

Results: We found no significant differences between patients' perceptions of effective communication using either IT-mediated communication or FtF communication with their physicians. However, we found significant differences in patients' perception of media effectiveness: patients perceived FtF communication to be a more favorable medium (P=.02). Interestingly, we found no significant difference in terms of benefits (P=.09) and success (P=.08) of IT-mediated communication versus FtF communication.

Conclusions: The results of this study imply that patients can achieve the same level of communication effectiveness with their physicians using IT-mediated communication as they would in comparable FtF interactions, but patients view FtF communication to be a more favorable medium than IT-mediated communication.

(J Med Internet Res 2020;22(5):e16965) doi:10.2196/16965

KEYWORDS

telemedicine; telemedicine; eHealth; text telecommunication; health communication

Introduction

Effective Communication Between Patients and Physicians

Effective communication with physicians is especially important for all patients and specifically for patients with chronic diseases, as these conditions require long-term continuous care [1]. Studies have shown that effective communication facilitates improvements in patients' physical health outcomes, self-care efficacy, and self-care management [2]. According to the Centers for Disease Control and Prevention, 60% of adults in the United States suffer from at least one chronic disease and 40% suffer from two or more. These patients account for 90%

of the US $3.3 trillion annual health care costs. Managing the symptoms of chronic disease can help to reduce these costs significantly.

In recent years, patients' and providers' use of technology has been gaining more prevalence as patients communicate more regularly with their physicians using technology [3] and an emphasis has been placed on telemedicine. According to a report by the American College of Surgeons [4], physicians no longer make house calls, and for a growing number of patients, text messaging and telemedicine have become alternatives to phone calls and traditional office visits. Several studies have shown patients' use of email, patient portals, and secure messaging to communicate with their physicians [5]. Research has also examined how information technology (IT)–mediated communication among patients with chronic disease can improve their health outcomes [6,7] and how media differ in effectiveness according to the communication process for which they are used (eg, scheduling an appointment and discussing acute symptoms). However, few studies have examined if patients' perceptions of communication effectiveness differ between IT-mediated communication and face-to-face (FtF) communication. Existing research mainly examined differences between FtF communication and video and voice communication [8-10]. Research about patients' perceptions of text-based IT-mediated communication is limited. Therefore, it is unclear if patients perceive the quality of care that is delivered FtF differently from text-based communication. As text-based IT-mediated communication is gaining more popularity [11], it is imperative to investigate how it impacts effective patient-physician communication. This gap in the literature creates an opportunity for health IT scholars to understand patients' perspectives regarding the effectiveness of text-based IT-mediated communication. Thus, we examine if patients' perceptions of communication effectiveness differ for FtF communication versus text-based IT-mediated communication.

Effective patient-physician communication is also an important factor in patient satisfaction, self-care efficacy, and perceived health management outcomes. Patients who experience effective communication with providers report greater satisfaction [12], self-care efficacy [13], and perceived health management outcomes [14]. These outcomes of effective communication are important, as patients who manage their health can improve their quality of life and reduce health care costs. Therefore, we also examine how perceptions of communication effectiveness facilitate patient satisfaction, self-care efficacy, and health management outcomes.

To date, researchers have studied the design, adoption, and use of information systems in health care extensively from the organizational and health provider perspective. However, there is paucity in understanding how the use of these emerging technologies can be compared with the use of traditional FtF models of care among patients. As effective communication can improve the quality of care for patients by empowering their self-care efficacy and health management, it is critical to understand patients' perceptions of effective communication using different channels and how it impacts their satisfaction, self-care efficacy, and health management outcomes. This study begins by reviewing relevant communication theory and health communication literature to understand the role that media play in patients' perceptions of effective communication. Next, it presents 2 competing models of effective communication, followed by the details and results of a national cross-sectional survey that tested them.

Background

IT-mediated communication has the potential to improve cost-effectiveness, quality, and accessibility of health care services [15-18]. Several studies have reported on telemedicine and telehealth in terms of its acceptance [19,20], adoption [21-23], compliance [24], quality [25], and trust [26]. Yet, it is unclear how the quality of patient-physician interactions differs between traditional FtF interactions and text-based interactions. Most research that has examined differences in FtF interactions and IT-mediated interactions has compared video and phone consultations with FtF consultations among healthy patients. For instance, research has shown that simulated patients who either had video or FtF consultations with medical interns did not experience differences in satisfaction, perceived information exchange, interpersonal relationship building, or perceived shared decision making [9]. Similarly, patients who received an FtF, a phone, or a video consultation reported no differences in distress levels after the consultation [27]. Moreover, small samples of patients who were screened for neurocognitive problems (N=8) or took part in cognitive interventions for the elderly (N=11) using video consultations or FtF consultations experienced no differences in cognitive or diagnostic outcomes [28,29], respectively, although communication effectiveness was not assessed. The literature is promising and suggests that there are no significant differences in the quality of care that is delivered using FtF, video, or phone consultations. At the same time, it is unclear if the same results would hold when comparing text-based channels and FtF communication among patients with chronic diseases who require regular communication with their providers and use text-based communication with their providers [3].

Cues-Filtered-In and Cues-Filtered-Out Perspectives

Frameworks that are useful for understanding the differences in communication effectiveness between FtF communication and IT-mediated communication are the cues-filtered-out and the cues-filtered-in perspectives. The cues-filtered-out perspective [30] assumes that IT-mediated communication hinders effective communication because of its reduced nonverbal cues (gestures, smiles, pats on the back, nods to show attentiveness, and eye contact), in comparison with channels that offer the transmission of more nonverbal cues such as the phone, videoconferencing, and FtF communication. A systematic review has shown the importance of nonverbal cues in patient-provider interaction in that when doctors make appropriate eye contact, do not interrupt patients, or pay attention to patients' nonverbal signals, patients' objective (eg, blood pressure) and subjective (eg, pain scores) health care outcomes improve [2]. The cues-filtered-out perspective echoes concerns about the use of telemedicine, such as the depersonalization of care, lack of physical presence, inhibition

of patient participation, and physician dominance of the medical encounter [31].

One theory that falls in the cues-filtered-out perspective is the social presence theory [32]. This theory posits that media differ in their capacity to transmit nonverbal and verbal information; therefore, the less cues a medium can transmit, the less warmth and involvement patients and providers experience with one another. Indeed, a systematic review about patients' experiences with remote monitoring for chronic conditions showed that patients view remote monitoring as jeopardizing interpersonal connections with their providers and do not want remote monitoring to replace FtF interactions [33]. Similarly, patients have reported missing FtF interaction with their providers when they receive telemedicine [34]. The abovementioned research suggests that patients experience less connection with their providers via telemedicine and do not want telemedicine to replace FtF communication. These results are in line with the cues-filtered-out perspective. Research has yet to show these same findings among text-based IT-mediated communication.

The cues-filtered-out perspective is in contrast to the cues-filtered-in perspective. In light of inconsistent findings for the cues-filtered-out perspective, a competing theory of social information processing was developed [35], thus creating the cues-filtered-in perspective. The cues-filtered-in perspective assumes that people adapt to the medium to achieve effective communication, regardless of the number of cues a medium transmits. This theory posits that patients and providers using IT-mediated communication can achieve the same outcomes as FtF communication, if interaction time is not restricted. Per the social information processing theory, patients and providers will use more verbal cues to achieve quality interactions and to exchange the same amount of information as they would if they were communicating FtF. In line with the social information processing theory, research has shown that simulated patients who used video, phone, or FtF communication with medical students in medical consultations reported no differences in patient satisfaction, information exchange, interpersonal relationship building, or shared decision making [9]. Similarly, patients who used either FtF communication or a video consultation with their physicians reported no differences in their physician's ability to develop rapport, use shared decision making, and/or promote patient-centered communication [31]. The abovementioned research suggests that patients do not perceive significant differences in communication using the phone, video, or FtF medical consultations, which is in line with the cues-filtered-in perspective. At the same time, research has yet to show this empirically among text-based IT-mediated communication.

Communication Effectiveness, Satisfaction, Self-Care Efficacy, and Health Management Outcomes

In addition to examining differences in patients' perceptions of communication effectiveness, this study also examines differences in patient satisfaction, self-care efficacy, and health management. Effective patient-physician communication is an important factor in patient satisfaction. A large-scale intervention has shown that communication skills training for 1537 physicians improved patients' satisfaction with provider communication [12]. At the same time, a systematic review of studies about patient satisfaction with interactive video consultations revealed mixed feelings about video consultations: Patients appreciate the accessibility of expert care, less travel, and reduced waiting times, but patients do not like communicating with their provider using video consultations [36]. Notably, this result is the opposite of the aforementioned studies that found that patients reported no differences in patient satisfaction when they used video, phone, or FtF communication in medical consultations with providers [9,31]. This research aimed to address these mixed results and advance the understanding regarding patients' perceptions of satisfaction.

Communication effectiveness is also important for self-care efficacy and health management. Patients who experience effective communication with providers (eg, feel listened to, respected, and that their provider explains things clearly) are more motivated to take care of themselves. Similarly, patients who have quality relationships with their providers (eg, "My provider listens carefully to me" and "It is easy to communicate with my provider") report greater levels of self-care efficacy and health care management [2,13]. Indeed, patients who have diabetes and effective provider communication report greater insulin adherence [37] and glycemic control [14]. We investigated the impact of communication effectiveness on patient satisfaction, self-care efficacy, and health management.

Methods

Research Model and Construct Development

To examine the differences in patients' perceptions of FtF communication and IT-mediated communication, we apply 2 competing perspectives to advance the understanding regarding the impact of patients' perceptions of effective communication and its impact on patient satisfaction, self-care efficacy, and health management outcomes. Per the cues-filtered-in perspective, patients should experience no differences in effective communication using IT-mediated communication (vs FtF communication) with their physicians, and, in turn, patients should report no differences in positive relationships between effective communication and self-care efficacy (H1A=H1B), effective communication and patient satisfaction (H2A=H2B), and effective communication and health management outcomes (H3A=H3B) using IT-mediated communication (vs FtF communication) with their physicians. Conversely, per the cues-filtered-out perspective, patients should experience more effective communication using FtF communication (vs IT-mediated communication) with their physicians, and, in turn, patients should report stronger positive relationships between effective communication and self-care efficacy (H1A<H1B), effective communication and patient satisfaction (H2A<H2B), and effective communication and health management outcomes (H3A<H3B) using FtF communication (vs IT-mediated communication) with their physicians.

Another important aspect of effective communication with providers is patients' perception of media effectiveness. Media effectiveness refers to the degree to which patients perceive that a specific medium helps them accomplish their communication

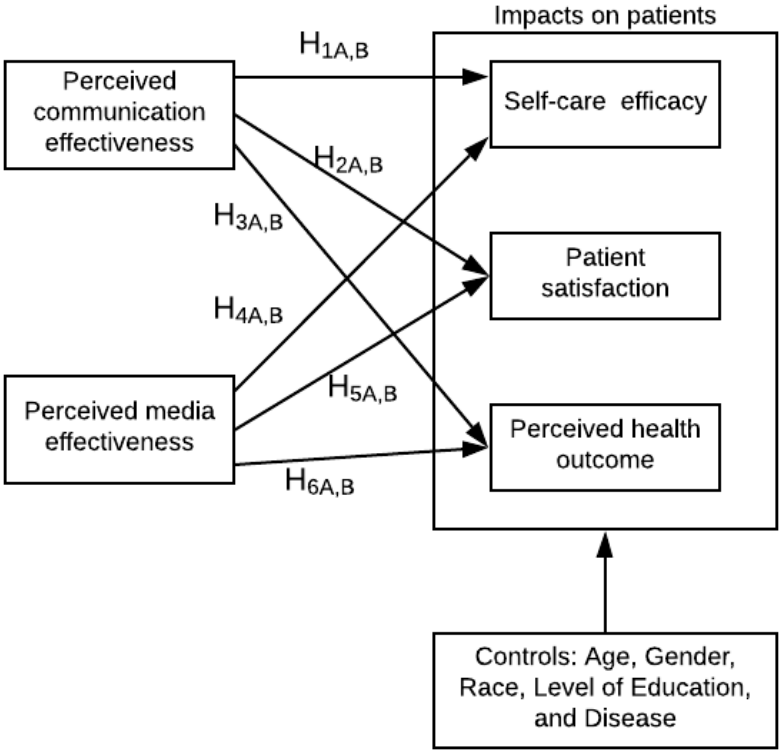
goal. Media effectiveness is operationalized using the same scale as communication effectiveness (eg, unsuccessful to successful, inefficient to efficient, and inappropriate to appropriate). However, we rephrased the questions to ask how effective the mode of communication is (media effectiveness), rather than how effective the interaction with one's provider is (communication effectiveness). As such, we are using this construct to separate *communication effectiveness with providers* from *communication effectiveness with a communication mode* to aid understanding about patients' perceptions of communication effectiveness with their providers using different media. Patients might describe interactions with their physicians as effective because they accomplish their communication goals, but they might not describe the mode of communication they used with their physician as effective because they might have preferred to use another mode of communication. Thus, we are attempting to differentiate between the different aspects of patients' perceived communication effectiveness (PCE) by parsing media effectiveness from communication effectiveness.

A similar concept to media effectiveness is media richness or a medium's capacity to facilitate shared understanding, as a rich medium facilitates insight and rapid understanding [38]. The media richness theory [39] operationalizes media richness based on objective characteristics of the medium, such as the speed of feedback, personal focus, the number of cues, and the ability to use natural language. As defined, media richness does not account for patients' perceptions of media effectiveness because it refers to intrinsic characteristics of a medium. As such, we are using the media effectiveness construct to understand any differences between patients' perceptions of communication effectiveness with providers, rather than the media richness construct.

The cues-filtered-out and cues-filtered-in perspectives offer 2 competing predictions for differences in perceived media effectiveness in physician interactions. Per the cues-filtered-in framework, patients should experience no differences in media effectiveness using IT-mediated communication (vs FtF communication) with their physicians, and, in turn, patients should report no differences in positive relationships between media effectiveness and self-care efficacy (H4A=H4B), media effectiveness and patient satisfaction (H5A=H5B), and media effectiveness and health management outcomes (H6A=H6B) using IT-mediated communication (vs FtF communication) with their physicians. Conversely, per the cues-filtered-out framework, patients should experience greater media effectiveness using FtF communication (vs IT-mediated communication) with their physicians, and, in turn, patients should report stronger positive relationships between media effectiveness and self-care efficacy (H4A<H4B), media effectiveness and patient satisfaction (H5A<H5B), and media effectiveness and health management outcomes (H6A<H6B) using FtF communication (vs IT-mediated communication) with their physicians.

To investigate these competing predictions, we first test the proposed model (Figure 1) separately for text-based IT-mediated communication (group A) and FtF communication (group B). Then, we compare the coefficients of PCE and perceived media effectiveness between the 2 groups in the Results section.

Figure 1. Effective communication comparison (face-to-face communication vs text-based information technology–mediated communication).



Data Collection

We designed and used a web-based questionnaire to survey a national sample of patients. To collect the data, we hired Qualtrics Panels. Qualtrics Panels is a market research company in the United States that partners with more than 20 sample providers to supply a network of diverse, quality respondents across the country. We used quota-based sampling to recruit approximately equal sample sizes of participants who communicated with their physician using text-based IT communication and FtF communication. Eligibility criteria included being aged at least 18 years, being located in the United States, and having engaged in communication with their physician in the previous month. Eligible participants in Qualtrics' panel were sent an incentivized invitation with the study link in various ways, such as email, a panel portal, and text messages. Participants were first presented with an informed consent document. After agreeing to participate, they began the survey questions. Qualtrics determines the monetary incentives for taking the survey, such as cash, gift cards, and vouchers. To avoid self-selection bias, survey invitations do not include specific details about the content of the survey and are kept very general. We ensured to use neutral wordings to develop the survey and used the semantic scale to avoid any common method bias. We found that none of the correlation values among the constructs exceed 0.90 [40]. Correlation matrix is provided in Multimedia Appendix 1.

Results

Characteristics of Participants

In total, 349 participants completed the survey, from which we used 345 complete usable responses. Participants' ages ranged from 35 to 85 years (mean 63, SD 9), and 45.2% (156/345) participants were male. A screening requirement of the survey was that all participants have been regularly communicating with their primary care physician over the past month. Of the 345 participants, 86 (24.9%) reported having diabetes, 69 (20.0%) reported arthritis, 52 (15.0%) reported heart disease, 34 (9.8%) reported pulmonary diseases, 28 (8.1%) reported psychiatric disorder, 17 (4.9%) reported hypertension, 17 (4.9%) reported chronic inflammatory disease, 16 (4.6%) reported thyroid, 12 (3.5%) reported immune system disorder, 7 (2.0%) reported hypertension, and 7 (2.0%) reported renal disease. For the most frequent mode of communication with physicians, of the 345, participants 134 (38.8%) reported FtF communication, whereas 211 (61.1%) participants reported the use of technology, including 152 (44.0%) patient portals, 121 (35.0%) email, and 72 (20.8%) text messaging. Of the 345 respondents, 45 (13.0%) participants completed a 2-year degree, 69 (20.0%) participants completed a 4-year degree, 72 (20.8%) participants earned a professional degree, and 145 (42.0%) participants attended some college.

Measurement Model

We used R software (R Project for Statistical Computing) and the Lavaan package to analyze the data. To test the proposed model (Figure 1), we measured PCE using 11 bipolar semantic differential scales [41]. Example items include insufficient/sufficient, adverse/beneficial, inadequate/adequate, unsuccessful/successful, useless/useful, disadvantageous/advantageous, and inefficient/efficient. We used the same scales to measure perceived media effectiveness. For the outcome variables including patient satisfaction, self-care efficacy, and perceived health management outcomes, we adapted survey questions from previous studies and used a semantic scale ranging from 1 (strongly disagree) to 5 (strongly agree) to measure each construct. The measurement properties of each construct are reported in Table 1.

Table 1. Psychometric properties of measurement model.

Constructs	Description	Number of items	Cronbach alpha	Average variance extracted	Source
Perceived media effectiveness	Patients' perception of medium's capacity to facilitate shared understanding	10	.94	0.83	Spitzberg [41]
Perceived communication effectiveness	The degree to which patients accomplished their communication goal	10	.95	0.84	Spitzberg [41]
Patient satisfaction	The degree to which patients experienced interest, accomplishment, and overall satisfaction in conversation with their provider	5	.93	0.82	Hecht [42]
Self-care efficacy	The degree to which the patients feel confident in managing their own care	7	.87	0.70	Plotnikoff et al [43]
Perceived health outcome	Patients' expectations of positive physical and self-evaluative outcomes	4	.91	0.72	Anderson et al [44]

We compared the proposed model for 2 groups based on the respondent's most frequent mode of communication with their physicians: group A (IT-mediated communication) versus group B (FtF communication). Following the common criteria

suggested in the literature [45,46], we examined the composite reliabilities of each construct and average variance extracted to ensure that the instrument had adequate reliability and convergent validity. Table 1 presents these statistics.

Hypothesis Testing

To evaluate the model's fit, we developed a structural equation model and tested for multigroup comparison while controlling for age, gender, race, the level of education, and the type of disease of the respondents. The results are presented in Table 2. To compare the 2 groups, we calculated a separate model for each group. If we assume that the residual values from the 2 models are normally distributed, we can test the hypothesis that the coefficient of PCE in group A is equal to the same coefficient in group B (ie, $beta_{PCE(A)}=beta_{PCE(B)}$) versus the alternative that they are unequal. The test statistic follows the Student t distribution [44] with v degrees of freedom where $v=n_1+n_2-4$ (n_1: sample size of group A and n_2: sample size of group B). The results of the group comparison analysis are presented in Table 3.

Table 2. Results for information technology–mediated communication and face-to-face communication groups.

Dependent and independent variables	Group A: information technology–mediated communication			Group B: face-to-face communication		
	Coefficient	SE	*P* value	Coefficient	SE	*P* value
Self-care efficacy						
Perceived communication effectiveness (H1)	0.116	0.025	<.001	0.167	0.029	<.001
Perceived media effectiveness (H4)	0.079	0.023	.001	0.158	0.026	<.001
Age	0.060	0.033	.07	0.312	0.667	.64
Gender	0.255	0.597	.67	0.036	0.038	.35
Level of education	0.450	0.286	.12	−0.128	0.267	.63
Patient satisfaction						
Perceived communication effectiveness (H2)	0.236	0.016	<.001	0.349	0.025	<.001
Perceived media effectiveness (H5)	0.166	0.017	<.001	0.281	0.027	<.001
Age	0.047	0.021	.03	0.062	0.033	.06
Gender	−0.35	0.383	.36	0.588	0.577	.31
Level of education	0.309	0.184	.09	0.595	0.231	.01
Perceived health outcome						
Perceived communication effectiveness (H3)	0.126	0.017	<.001	0.171	0.025	<.001
Perceived media effectiveness (H6)	0.078	0.016	<.001	0.139	0.024	<.001
Age	−0.033	0.022	.14	0.041	0.033	.21
Gender	0.275	0.403	.49	0.758	0.576	.19
Level of education	0.236	0.193	.23	0.19	0.23	.41

Table 3. Model comparison across groups.

Hypotheses	*t* value (*df*)	*P* value
H1A[a] versus H1B[b]	1.362 (333)	.17
H2A versus H2B	3.859 (333)	<.001
H3A versus H3B	1.518 (333)	.13
H4A versus H4B	2.261 (333)	.02
H5A versus H5B	3.578 (333)	<.001
H6A versus H6B	2.093 (333)	.04

[a]A: information technology–mediated group.

[b]B: face-to-face group.

The results showed that among both groups, PCE and perceived media effectiveness had a significant positive impact on patients' self-efficacy (H1A, H1B, H4A and H4B), patient satisfaction (H2A, H2B, H5A and H5B), and perceived health management outcomes (H3A, H3B, H6A and H6B). The data also revealed that there is no significant difference in PCE between FtF communication and IT-mediated communication, yet there is a significant difference between the 2 groups in terms of perceived media effectiveness. Patients perceive FtF

communication to be more effective in improving their self-efficacy than IT-mediated communication.

Additional results found that both PCE and media effectiveness have a significant impact on improving patients' satisfaction. There is also a significant difference in the impact of PCE and media effectiveness on patient satisfaction. The impact of communication effectiveness and media effectiveness on patient satisfaction is significantly higher for FtF communication in comparison with IT-mediated communication.

Furthermore, the results revealed that PCE and media effectiveness significantly improved patients' perceived health outcomes. Interestingly, there is no significant difference in the impact of PCE on perceived health outcomes across groups. However, there is a significant difference in the impact of perceived media effectiveness on perceived health outcomes. Patients who communicate FtF revealed a significantly higher association between media effectiveness and perceived health outcomes.

To further understand the differences in PCE and perceived media effectiveness, we compared the IT-mediated group A's and the FtF communication group B's responses for each scale item using *t* tests. We found no significant difference between FtF communication and IT-mediated communication groups in items that form PCE. We did, however, find significant differences between FtF communication and IT-mediated communication groups in items that formed perceived media effectiveness. There were significant differences in perceptions of media effectiveness in terms of sufficiency, adequacy, advantage, favorableness, and suitability. Patients who primarily used FtF communication (vs IT-mediated communication) assigned significantly higher scores to the abovementioned items but assigned no significant difference between the 2 groups in terms of media appropriateness, benefit, success, usefulness, and efficiency. The test results are presented in Table 4.

Table 4. Group comparison for perceived communication effectiveness and perceived media effectiveness.

Semantic scale	Perceived communication effectiveness		Perceived media effectiveness	
	t value (*df*)	*P* value	*t* value (*df*)	*P* value
Inappropriate: appropriate	−0.53 (333)	.598	−1.24 (333)	.21
Insufficient: sufficient	−0.25 (333)	.80	−2.62 (333)	.01
Adverse: beneficial	−0.97 (333)	.33	−1.70 (333)	.09
Inadequate: adequate	−1.11 (333)	.27	−3.23 (333)	.001
Unsuccessful: successful	−0.63 (333)	.53	−1.74 (333)	.08
Useless: useful	−1.07 (333)	.29	−1.76 (333)	.08
Disadvantageous: advantageous	−0.03 (333)	.98	−2.47 (333)	.01
Unfavorable: favorable	−1.18 (333)	.24	−2.43 (333)	.01
Inefficient: efficient	−0.50 (333)	.62	−1.47 (333)	.14
Unsuitable: suitable	−0.91 (333)	.37	−2.58 (333)	.01

Discussion

Theoretical Implications

In this study, we applied competing approaches to advance the understanding regarding patients' perceptions of effective communication with their primary care physicians. In a national sample of patients with chronic diseases, we compared patients' perceptions of effective communication with their provider using either text-based IT-mediated communication (email, patient portal, and messaging) or FtF communication. The primary results revealed no significant differences between patients' perceptions of effective communication using either IT-mediated communication or FtF communication, which is in line with the cues-filtered-in perspective. Interestingly, at a more granular level, patients perceived FtF communication to be a more favorable medium than IT-mediated communication, which is in line with the cues-filtered-out perspective. As a result, differences in perceived media effectiveness impacted differences in patients' self-care efficacy, satisfaction, and perceived health management outcomes between patients who used either FtF communication or IT-mediated communication with their physicians. The results imply that patients can achieve the same level of communication effectiveness with their physicians using IT-mediated communication as they would in comparable FtF interactions, but patients view FtF communication to be a more favorable medium than IT-mediated communication.

The results are promising as they show that as patients increasingly use different forms of IT-mediated communication, such as patient portals, email, and text messaging, to communicate with their physicians, patients will continue to adapt to the medium and achieve the same level of effectiveness, satisfaction, self-care efficacy, and health care outcomes as they would in comparable FtF interactions with their physicians. At the same time, the data show that although patients report no differences in effective communication with their physicians using IT-mediated communication or FtF communication, patients report FtF communication as a more effective mode of communication relative to IT-mediated communication. In turn, the positive relationships between media effectiveness and self-care efficacy, patient satisfaction, and perceived health

management were stronger for patients who used FtF communication.

Practical Implications

The results of this study are consistent with previous research that has shown that patients generally describe their experience with telehealth positively, yet they still value their FtF contact with health care professionals [32,33]. Patients with chronic diseases have reported many benefits of IT-mediated communication, such as improved self-management, shared decision making, better access to health care, and peace of mind. However, these potential benefits are often balanced against concerns about losing interpersonal contact. Patients often view IT-mediated communication as jeopardizing interpersonal connections with their providers and an unsuitable replacement for FtF interactions. These concerns are evident in the data, as further exploration of the underlying items that led to differences in perceived media effectiveness between FtF communication and IT-mediated communication showed that patients perceived FtF communication as a significantly more sufficient, adequate, advantageous, and favorable medium than IT-mediated communication. Notably, there were no other significant differences between FtF communication and IT-mediated communication in semantic scales that patients used to describe media effectiveness such as usefulness, success, appropriateness, and beneficial. These results suggest that patients perceive IT-mediated communication to be an effective medium for accomplishing communication goals with physicians (eg, share blood pressure data and make an appointment), but they prefer FtF communication and view it as a more suitable and favorable medium.

These mixed findings often lead researchers to infer that the sustained use of IT-mediated communication will be ensured with occasional FtF visits to physicians [47]. Nonetheless, these findings suggest a fundamental difference in how researchers and practitioners should approach patients' adoption and use of telemedicine. More research needs to be done to shift the focus from examining the usefulness of the technology to examining how best to educate users to use the technology to improve relationships with physicians. Telemedicine holds great potential for reducing the variability of diagnoses as well as improving clinical management and delivery of health care services by enhancing access, quality, efficiency, and cost-effectiveness [48,49]. Furthermore, evidence points to important socioeconomic benefits to patients, families, health practitioners, and the health system, including enhanced patient-physician communication opportunities [50]. IT-mediated communication has yet to be consistently employed in the health care system to deliver routine services. Creating effective relationships between patients with chronic diseases and physicians is important. When patients and physicians work together to determine optimal treatment plans in a value-centered manner, this significantly improves patients' confidence in self-care, satisfaction, and self-care management. These improvements can happen through text-based IT-mediated communication [24]. However, there is a need to educate patients about the effectiveness of IT-mediated communication. Managerial multidisciplinary efforts that draw expertise from communication sciences, health informatics and IT, public health, and health management and policy are required to ensure that telemedicine and electronic health systems are designed with a patient-centered focus and with attention to educating patients about how to use text-based technology to communicate with their physicians. Patient-physician communication will continue to evolve with time, both as a byproduct of technological advances as well as shifting societal values. In urban or developed areas, hospital beds are mainly occupied by patients with chronic diseases, decreasing bed availability for other patients in need. Recently, Centers for Medicare and Medicaid Services published a proposal for the management of patients with chronic illness that would allow physicians to be paid for non-FtF encounters. The realization of these innovative initiatives calls for changing the provider culture and workflow systems to allow the full incorporation of telemedicine into traditional care. Furthermore, patients need to be educated about how to interact with and use these technologies to communicate effectively with their providers and to manage their own care.

Conclusions

In this study, we investigated differences in patients' perceptions of IT-mediated communication and FtF communication. The results support that there is no significant difference in effective communication for using technology versus FtF communication. However, patients perceive FtF communication as a significantly more favorable, suitable, and sufficient medium in comparison with IT-mediated communication. More research about how to educate patients to use technology and how to use technology to improve relationships with physicians is needed.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Correlations matrix.
[DOCX File , 21 KB - jmir_v22i5e16965_app1.docx]

References

1. World Health Organization. Noncommunicable Diseases Country Profiles 2018. Geneva: World Health Organization; 2018.
2. Kelley JM, Kraft-Todd G, Schapira L, Kossowsky J, Riess H. The influence of the patient-clinician relationship on healthcare outcomes: a systematic review and meta-analysis of randomized controlled trials. PLoS One 2014;9(4):e94207 [FREE Full text] [doi: 10.1371/journal.pone.0094207] [Medline: 24718585]

3. 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]
4. Shipper ES, Hardaway JC, Garvey EM, Logghe H. Talking through time: trends in communication and the evolving patient-physician relationship. Bull Am Coll Surg 2016 Aug;101(8):19-23. [Medline: 28941427]
5. Lee JL, Choudhry NK, Wu AW, Matlin OS, Brennan TA, Shrank WH. Patient use of email, Facebook, and physician websites to communicate with physicians: a national online survey of retail pharmacy users. J Gen Intern Med 2016 Jan;31(1):45-51 [FREE Full text] [doi: 10.1007/s11606-015-3374-7] [Medline: 26105675]
6. Knox L, Rahman RJ, Beedie C. Quality of life in patients receiving telemedicine enhanced chronic heart failure disease management: A meta-analysis. J Telemed Telecare 2017 Aug;23(7):639-649. [doi: 10.1177/1357633X16660418] [Medline: 27450573]
7. Vatnøy TK, Thygesen E, Dale B. Telemedicine to support coping resources in home-living patients diagnosed with chronic obstructive pulmonary disease: patients' experiences. J Telemed Telecare 2017 Jan;23(1):126-132. [doi: 10.1177/1357633X15626854] [Medline: 26853926]
8. 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]
9. Tates K, Antheunis ML, Kanters S, Nieboer TE, Gerritse MB. The effect of screen-to-screen versus face-to-face consultation on doctor-patient communication: an experimental study with simulated patients. J Med Internet Res 2017 Dec 20;19(12):e421 [FREE Full text] [doi: 10.2196/jmir.8033] [Medline: 29263017]
10. Zahedi FM, Walia N, Jain H. Augmented virtual doctor office: theory-based design and assessment. J Manag Inf Syst 2016;33(3):776-808. [doi: 10.1080/07421222.2016.1243952]
11. Horwitz LI, Detsky AS. Physician communication in the 21st century: to talk or to text? J Am Med Assoc 2011 Mar 16;305(11):1128-1129. [doi: 10.1001/jama.2011.324] [Medline: 21406650]
12. Boissy A, Windover AK, Bokar D, Karafa M, Neuendorf K, Frankel RM, et al. Communication skills training for physicians improves patient satisfaction. J Gen Intern Med 2016 Jul;31(7):755-761 [FREE Full text] [doi: 10.1007/s11606-016-3597-2] [Medline: 26921153]
13. Eton DT, Ridgeway JL, Linzer M, Boehm DH, Rogers EA, Yost KJ, et al. Healthcare provider relational quality is associated with better self-management and less treatment burden in people with multiple chronic conditions. Patient Prefer Adherence 2017;11:1635-1646 [FREE Full text] [doi: 10.2147/PPA.S145942] [Medline: 29033551]
14. Linetzky B, Jiang D, Funnell MM, Curtis BH, Polonsky WH. Exploring the role of the patient-physician relationship on insulin adherence and clinical outcomes in type 2 diabetes: Insights from the MOSAIc study. J Diabetes 2017 Jun;9(6):596-605. [doi: 10.1111/1753-0407.12443] [Medline: 27368146]
15. Chiasson MW, Davidson E. Pushing the contextual envelope: developing and diffusing IS theory for health information systems research. Inform Organ 2004;14(3):155-188. [doi: 10.1016/j.infoandorg.2004.02.001]
16. Devaraj S, Kohli R. Information technology payoff in the health-care industry: a longitudinal study. J Manag Inf Syst 2000;16(4):41-67. [doi: 10.1080/07421222.2000.11518265]
17. McCullough JS, Casey M, Moscovice I, Prasad S. The effect of health information technology on quality in US hospitals. Health Aff (Millwood) 2010 Apr;29(4):647-654. [doi: 10.1377/hlthaff.2010.0155] [Medline: 20368594]
18. Schoen C, Davis K, How SK, Schoenbaum SC. US health system performance: a national scorecard. Health Aff (Millwood) 2006;25(6):w457-w475. [doi: 10.1377/hlthaff.25.w457] [Medline: 16987933]
19. Chau PY, Hu PJ. Investigating healthcare professionals' decisions to accept telemedicine technology: an empirical test of competing theories. Inform Manag 2002;39(4):297-311. [doi: 10.1016/s0378-7206(01)00098-2]
20. Hu PJ, Chau PY, Sheng OR, Tam KY. Examining the technology acceptance model using physician acceptance of telemedicine technology. J Manag Inf Syst 1999;16(2):91-112. [doi: 10.1080/07421222.1999.11518247]
21. Cho S, Mathiassen L, Robey D. Dialectics of resilience: a multi–level analysis of a telehealth innovation. J Inf Technol 2007 Mar;22(1):24-35. [doi: 10.1057/palgrave.jit.2000088]
22. LeRouge C, Garfield M. Crossing the telemedicine chasm: have the US barriers to widespread adoption of telemedicine been significantly reduced? Int J Environ Res Public Health 2013 Nov 28;10(12):6472-6484 [FREE Full text] [doi: 10.3390/ijerph10126472] [Medline: 24287864]
23. Rajan B, Tezcan T, Seidmann A. Service systems with heterogeneous customers: investigating the effect of telemedicine on chronic care. Manag Sci 2019 Mar;65(3):955-1453. [doi: 10.1287/mnsc.2017.2979]
24. Paré G, Jaana M, Sicotte C. Systematic review of home telemonitoring for chronic diseases: the evidence base. J Am Med Inform Assoc 2007;14(3):269-277. [doi: 10.1197/jamia.M2270] [Medline: 17329725]
25. LeRouge C, Hevner AR, Collins RW. It's more than just use: an exploration of telemedicine use quality. Decis Supp Syst 2007;43(4):1287-1304. [doi: 10.1016/j.dss.2006.02.007]
26. Paul DL, McDaniel Jr RR. A field study of the effect of interpersonal trust on virtual collaborative relationship performance. Manag Inf Syst Q 2004 Jun;28(2):183-227. [doi: 10.2307/25148633]

27. Idan A, Wallach HS, Almagor M, Waisman Y, Linn S. Mediated telemedicine vs face-to-face medicine: efficiency in distress reduction. J Multimodal User Interfaces 2015;9(4):333-339. [doi: 10.1007/s12193-015-0181-8]
28. Grosch MC, Weiner MF, Hynan LS, Shore J, Cullum CM. Video teleconference-based neurocognitive screening in geropsychiatry. Psychiatry Res 2015 Feb 28;225(3):734-735 [FREE Full text] [doi: 10.1016/j.psychres.2014.12.040] [Medline: 25596957]
29. Poon P, Hui E, Dai D, Kwok T, Woo J. Cognitive intervention for community-dwelling older persons with memory problems: telemedicine versus face-to-face treatment. Int J Geriatr Psychiatry 2005 Mar;20(3):285-286. [doi: 10.1002/gps.1282] [Medline: 15717335]
30. Culnan MJ, Markus ML. Information technologies. In: Jablin FM, Putnam LL, Roberts KH, Porter LW, editors. Handbook of Organizational Communication: An Interdisciplinary Perspective. Thousand Oaks, CA: SAGE Publications; 1987.
31. Agha Z, Schapira RM, Laud PW, McNutt G, Roter DL. Patient satisfaction with physician-patient communication during telemedicine. Telemed J E Health 2009 Nov;15(9):830-839. [doi: 10.1089/tmj.2009.0030] [Medline: 19919189]
32. Short J, Williams E, Christie B. The Social Psychology of Telecommunications. Hoboken, New Jersey: John Wiley & Sons; 1976.
33. Walker RC, Tong A, Howard K, Palmer SC. Patient expectations and experiences of remote monitoring for chronic diseases: systematic review and thematic synthesis of qualitative studies. Int J Med Inform 2019 Apr;124:78-85 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.01.013] [Medline: 30784430]
34. Gorst SL, Coates E, Armitage CJ. 'It's sort of a lifeline': chronic obstructive pulmonary disease patients' experiences of home telehealth. Health Psychol 2016 Jan;35(1):60-68. [doi: 10.1037/hea0000246] [Medline: 26148189]
35. Walther JB. Interpersonal effects in computer-mediated interaction. Commun Res 1992;19(1):52-90. [doi: 10.1177/009365092019001003]
36. Mair F, Whitten P. Systematic review of studies of patient satisfaction with telemedicine. Br Med J 2000 Jun 3;320(7248):1517-1520 [FREE Full text] [doi: 10.1136/bmj.320.7248.1517] [Medline: 10834899]
37. White RO, Eden S, Wallston KA, Kripalani S, Barto S, Shintani A, et al. Health communication, self-care, and treatment satisfaction among low-income diabetes patients in a public health setting. Patient Educ Couns 2015 Feb;98(2):144-149 [FREE Full text] [doi: 10.1016/j.pec.2014.10.019] [Medline: 25468393]
38. Daft RL, Lengel RH, Trevino LK. Message equivocality, media selection, and manager performance: Implications for information systems. Manag Inf Syst Q 1987 Sep;11(3):355. [doi: 10.2307/248682]
39. Dennis AR, Kinney ST. Testing media richness theory in the new media: the effects of cues, feedback, and task equivocality. Inf Syst Res 1998 Sep;9(3):256-274. [doi: 10.1287/isre.9.3.256]
40. Muthen LK, Muthen BO. Muthén & Muthén, Mplus. 2015. Mplus: Statistical Analysis With Latent Variables. User's Guide URL: https://www.statmodel.com/download/usersguide/MplusUserGuideVer_7.pdf [accessed 2019-05-05]
41. Spitzberg BH. Communication competence: measures of perceived effectiveness. In: Tardy CH, editor. Communication and Information Science. A Handbook for the Study of Human Communication: Methods and Instruments for Observing, Measuring, and Assessing Communication Processes. New York, NY: Ablex Publishing; 1988:67-105.
42. Hecht ML. Measures of communication satisfaction. Human Comm Res 1978;4(4):350-368. [doi: 10.1111/j.1468-2958.1978.tb00721.x]
43. Plotnikoff RC, Lippke S, Johnson ST, Courneya KS. Physical activity and stages of change: a longitudinal test in types 1 and 2 diabetes samples. Ann Behav Med 2010 Oct;40(2):138-149. [doi: 10.1007/s12160-010-9193-5] [Medline: 20533010]
44. Anderson ES, Winett RA, Wojcik JR. Self-regulation, self-efficacy, outcome expectations, and social support: social cognitive theory and nutrition behavior. Ann Behav Med 2007;34(3):304-312. [doi: 10.1007/bf02874555] [Medline: 18020940]
45. Chin WW. How to write up and report PLS analyses. In: Esposito VV, Chin W, Henseler J, Wang H, editors. Handbook of Partial Least Squares. Berlin, Heidelberg: Springer; 2010:655-690.
46. Henseler J, Ringle CM, Sinkovics RR. The use of partial least squares path modeling in international marketing. In: Sinkovics R, Ghauri P, editors. New Challenges to International Marketing. Bingley: Emerald Group Publishing Limited; 2009:277-319.
47. Rajan B, Seidmann A, Dorsey ER. The competitive business impact of using telemedicine for the treatment of patients with chronic conditions. J Manag Inf Syst 2013;30(2):127-158. [doi: 10.2753/mis0742-1222300205]
48. Dupont WD, Plummer WD. Power and sample size calculations for studies involving linear regression. Control Clin Trials 1998 Dec;19(6):589-601. [doi: 10.1016/s0197-2456(98)00037-3] [Medline: 9875838]
49. Craig J, Patterson V. Introduction to the practice of telemedicine. J Telemed Telecare 2005;11(1):3-9. [doi: 10.1177/1357633X0501100102] [Medline: 15829036]
50. Jennett PA, Hall LA, Hailey D, Ohinmaa A, Anderson C, Thomas R, et al. The socio-economic impact of telehealth: a systematic review. J Telemed Telecare 2003;9(6):311-320. [doi: 10.1258/135763303771005207] [Medline: 14680514]

Abbreviations

FtF: face-to-face
IT: information technology

PCE: perceived communication effectiveness

Edited by G Eysenbach; submitted 07.11.19; peer-reviewed by F Meza, S Edirippulige; comments to author 21.01.20; revised version received 30.01.20; accepted 07.02.20; published 13.05.20.

Please cite as:
Mirzaei T, Kashian N
Revisiting Effective Communication Between Patients and Physicians: Cross-Sectional Questionnaire Study Comparing Text-Based Electronic Versus Face-to-Face Communication
J Med Internet Res 2020;22(5):e16965
URL: https://www.jmir.org/2020/5/e16965
doi:10.2196/16965
PMID:32401213

Original Paper

Patient Empowerment Using Electronic Telemonitoring With Telephone Support in the Transition to Insulin Therapy in Adults With Type 2 Diabetes: Observational, Pre-Post, Mixed Methods Study

Helen McGloin[1], PgDip, MSc, PhD; Dympna O'Connell[1], BNS, MSc; Michele Glacken[2], BSc, MSc, PhD; Patsy Mc Sharry[1], BA, MMSc, PhD; Denise Healy[1], PgDip, MSc; Lisa Winters-O'Donnell[1], BSc, PgDip, MSc; Kathleen Crerand[3], HDip, MSc; Anne Gavaghan[3], HDip, BSc; Louise Doherty[3], MPH, MA, MB

[1]Department of Nursing, Health Science and Disability Studies, St Angela's College, Sligo, Ireland

[2]Institute of Technology, Sligo, Ireland

[3]HSE West, Letterkenny, Donegal, Ireland

Corresponding Author:
Helen McGloin, PgDip, MSc, PhD
Department of Nursing, Health Science and Disability Studies
St Angela's College
Lough Gill
Sligo, F91C634
Ireland
Phone: 353 719135610
Email: hmcgloin@stangelas.nuigalway.ie

Abstract

Background: Initiation of insulin therapy for the management of type 2 diabetes can be an unwelcome and distressful development for patients. Current evidence suggests that telemonitoring can help improve glycemic control in type 2 diabetes and can support empowerment to self-manage diabetes. This telemonitoring intervention was underpinned by an empowerment approach.

Objective: This study aimed to evaluate the clinical effectiveness and feasibility and the patients' and health care providers' experiences of a 12-week telemonitoring intervention with telephone support for patients commencing insulin therapy. This paper focuses on the impact on patient empowerment.

Methods: An observational, pre-post, multimethod, and triangulation design was employed to study a 12-week automated electronic telemonitoring intervention with telephone support from a diabetes clinical nurse specialist (CNS). Forty patients were recruited from the clinic as they were about to commence insulin therapy. In the quantitative arm, biometric data (hemoglobin A_{1c} [HbA_{1c}] and weight) and psychosocial data (diabetes empowerment scale [DES] scores and diabetes distress scale [DDS] scores) were gathered by the research team at baseline (T1), the end of the intervention (T2), and 3 months postintervention (T3). Data on hospital admission and general practitioner (GP) visits were collected for the duration of the study. In the qualitative arm, separate focus group interviews were conducted with the CNS team supporting the intervention (n=2) and patients (n=16).

Results: Of 39 patients who completed the intervention, 23 (59%) were male. The mean age of the sample was 62.4 years (range 37-80 years). The mean HbA_{1c} (mmol/mol) decreased significantly between T1 and T2 (mean difference [MD] −17.13; *P*<.001) and T1 and T3 (MD −18.16; *P*<.001), with no significant impact on weight. In the focus groups, patients reported an increased awareness to self-manage diabetes and feelings of safety and comfort. There were 13% (5/39) of patients who had hypoglycemia on two or more occasions. A significant increase in the mean DES score occurred between T1 and T2 (MD 0.62; *P*=.001) and T1 and T3 (MD 0.72; *P*<.001). The mean DDS score decreased between T1 and T2 (MD −0.64; *P*=.002) and T1 and T3 (MD −0.6; *P*=.002). The mean patient satisfaction with the intervention was above 4 out of possible 5 on all items on the Telemedicine Satisfaction and Usefulness Questionnaire. We observed a reduction in diabetes clinic attendances and GP visits. A significant increase in workload was reported by the CNS team.

Conclusions: This intervention had an empowering effect for patients in the self-management of type 2 diabetes and has the potential to meet the need for safer and more effective care in insulin initiation in the community setting. We observed a significant

increase in workload for health care staff. Telemonitoring needs to be streamlined with health care delivery and accompanied by adequate support services.

(J Med Internet Res 2020;22(5):e16161) doi:10.2196/16161

KEYWORDS

diabetes mellitus; type 2; telehealth; insulin; empowerment; self-management

Introduction

Background

The goal of diabetes treatment is optimal glycemic control and prevention of complications. International guidelines recommend the initiation of insulin when people with type 2 diabetes have signs and symptoms of acute decompensation that are no longer controlled by oral hypoglycemic drugs and lifestyle [1]. Insulin initiation is often delayed, and once started, achieving optimal doses requires a frequent scheduled review of blood glucose levels and individualized dose titration [2]. The failure to commence and intensify treatment is termed *clinical inertia* and is a considerable problem in primary care practice [3]. There are multiple barriers to optimal insulin treatment in primary care at the patient and practitioner levels, and there is a need for interventions to improve self-management support and integrated insulin support systems [4].

Once insulin is commenced, many patients are not followed up until the next clinic appointment, resulting in delays in achieving the optimal insulin dosage [5]. The barriers to initiating insulin include misconceptions about insulin, perceived difficulty in management for both physicians and patients, and risk of hypoglycemia and weight gain [6]. Lack of time is a common health care system–related barrier to insulin initiation [5]. Starting insulin is stressful for patients, and providing adequate support and monitoring during this process can present challenges for health care providers, particularly in between health care visits [7]. The safe and effective transition to insulin therapy requires health care providers to examine alternative ways of empowering the patient in self-management.

Empowerment can be viewed as both a process and an outcome [8]. It can be viewed as a process when an intervention aims to equip patients and their families with the self-awareness, autonomy, knowledge, and skills to become *comanagers* of their condition in partnership with health professionals. The desired outcome is enhanced confidence and skills to manage the physical, emotional, and social impacts of diabetes in their daily lives. Patients who feel empowered by their health care practitioner develop a greater sense of self-control, which may lead to better glycemic control [9]. Positive effects of telemonitoring include patient empowerment to self-manage [10] healthy coping and problem solving [11], and improved practitioner-patient relationships and patient engagement have also been reported [10-13].

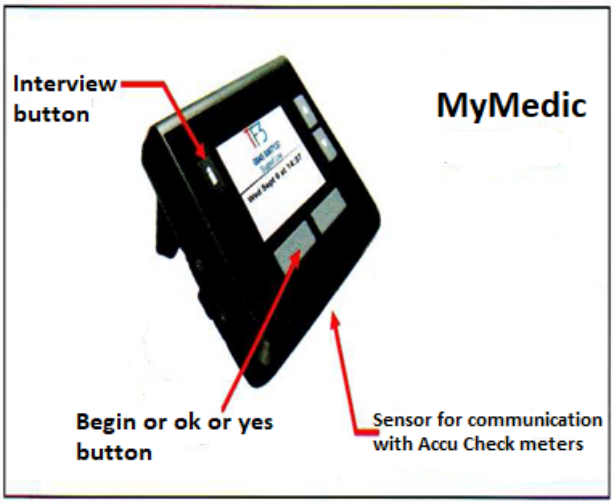
A review by Risling et al [14] found that the consistent lack of conceptual clarity on what constitutes empowerment in the electronic health (eHealth) context means that evaluating patient empowerment associated with eHealth technology is challenging. More recently, Risling et al [15] found that patients in a digital study identified the relational and informational elements of empowerment and recommended that these key areas should shape the focus of the evolution of patient empowerment in digital research. Telemonitoring has been found to be effective as a confidence, decision-making, and self-care enhancer in older persons with chronic heart failure [16,17]. Telemonitoring surveillance systems have left patients with feelings of greater self-control, heightened motivation for lifestyle changes, and improved quality of life [18].

Aims

This study sought to implement and evaluate telemonitoring with telephone support in a real-world setting using an observational, pre-post, mixed methods design in a cohort of patients who were about to commence insulin therapy. A MyMedic hub (Figure 1) was placed in the patients' home for 12 weeks. The patients recorded their blood glucose readings as normal, and at the times agreed, they sent their blood glucose readings to a monitoring center using the hub. The schedule for upload was agreed between the diabetes team and the patients, and the hub unit prompted the individual to upload their readings as agreed. The team reviewed the blood glucose results and contacted the participant, if necessary, to seek additional information on their symptoms and well-being. Insulin was adjusted according to need via telephone discussions using an insulin adjustment plan.

This paper aimed to describe the impact on hemoglobin A_{1c} (HbA_{1c}), hypoglycemic events, patient empowerment, diabetes distress, and satisfaction with telemonitoring from the patients' perspective. In addition, the experience of using telemonitoring to facilitate the transition to insulin therapy was explored from the perspective of the diabetes team and health care organization.

Figure 1. MyMedic hub.



Methods

Study Design

An observational, pre-post, multimethod, triangulation design was used to evaluate the feasibility, clinical effectiveness, and resource implications of telemonitoring support for patients commencing insulin therapy in the real-world setting. The design was underpinned by an empowerment philosophy. The mixed methods approach allowed for a holistic overview of using telemonitoring in diabetes care, generating physiological data and exploring receptiveness of technology and the core aspects of empowerment.

Population

To be eligible for inclusion in the study, patients had to be aged older than 18 years with type 2 diabetes and commencing with insulin therapy. Patients with a score below 4 on the clock drawing test were excluded, as this indicates the cognitive inability to self-manage insulin [19]. The patients were recruited by the clinical nurse specialist (CNS) team from hospitals and community-based diabetes clinics in the Northwest of Ireland. The sample size was determined by cost, available time of the CNS to support, telemonitoring equipment, and timeline for the feasibility study resources to provide the intervention. The project team determined that 40 patients would be enough to yield useful data and be achievable within the available resources to fund the project.

Sample

Convenience sampling was used. When the decision to commence a patient on insulin was made, the CNS invited them to participate in the study. One participant chose to discontinue from the study after 1 week because of the problems experienced with connectivity. The recruitment took place from April 2016 to June 2017.

Intervention

Following recruitment and consent, the patients were commenced on insulin therapy and provided with the telemonitoring system in addition to standard care. Standard care of this cohort includes telephone calls with the CNS with visits to the diabetes clinic as needed to collate blood glucose readings and to adjust insulin levels once a week.

A MyMedic hub (Figure 1) was placed in the patients' home for 12 weeks by a telecare support officer from Fold TeleCare who was contracted to provide the monitoring on behalf of the Health Service Executive (HSE), Ireland. The telecare support officer who installed the hub in the home taught the patients how to send their readings and gave instructions about what to do if anything did not work.

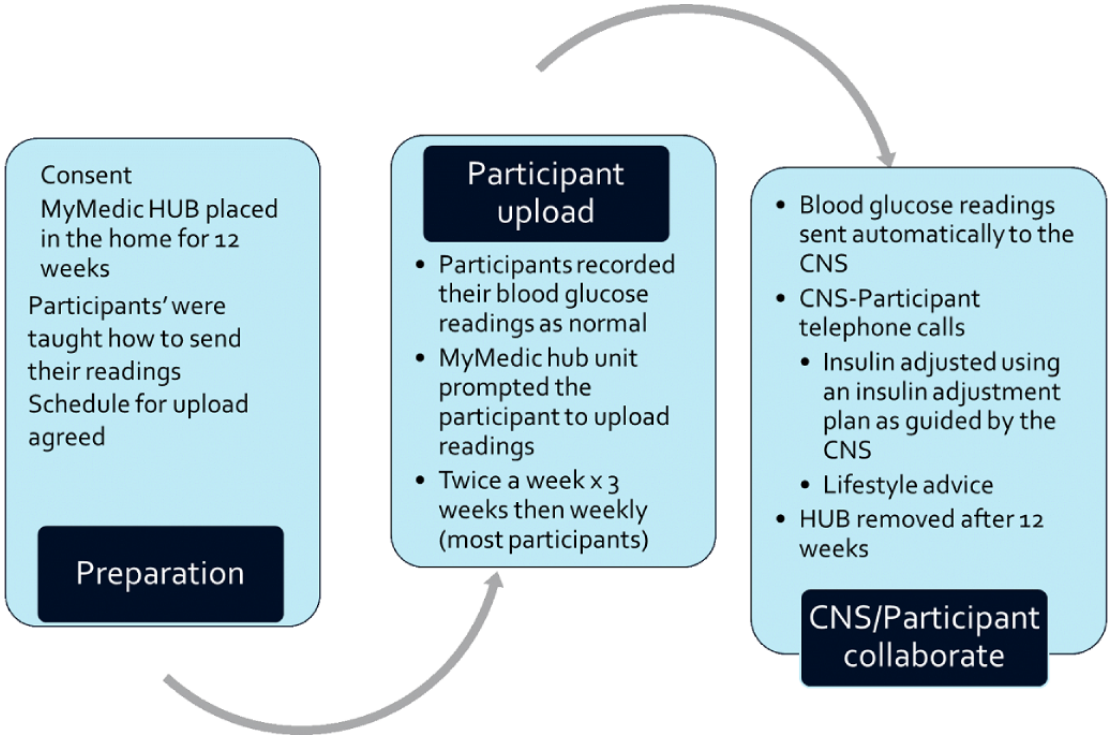
The patients recorded their blood glucose readings as instructed by the CNS. The usual practice is that patients on basal insulin self-test their blood glucose levels four times daily or more to establish the pattern of glucose through the day and to comply with the HSE guidelines [20]. At the times agreed, they sent their blood glucose readings to a monitoring center using the hub. The hub unit prompted patients to upload their readings as agreed with the CNS. Most patients uploaded twice a week for the first 3 weeks and then weekly for the remaining 9 weeks, as they needed more intensive support at the beginning. This was based on a similar study by Turner et al [21]. We contacted the authors to find out about their titration guide and frequency of dose adjustment, which was every 3 days initially and then weekly once stable. The American Diabetes Association and the European Association for the Study of Diabetes joint guidelines [22] recommend increasing basal insulin dose by 2 IU every 3 days until fasting glucose reaches the desired range individualized for each patient. As the target is neared, dosage adjustments should be more modest and occur less frequently.

None of the patients required more than weekly uploads after week 3. If the participants did not upload their blood glucose levels at the scheduled time, the telecare support officer contacted them by telephone to remind them.

The readings were sent automatically to the web-based platform. The CNS accessed the data for the individual on a prearranged date and contacted the participant, if necessary, to seek additional information on their symptoms and well-being. The record was displayed in a diary format, showing the patient's

readings at each time of the day. This allowed the review of glucose data throughout the day. Charts, graphs, and tables could be created from these data to allow analysis. Insulin doses were entered onto the system with any changes to the care plan to allow sharing of data between health professionals. Blood glucose data were saved onto the patients' health service electronic file in pdf. Patients could contact their CNS or general practitioner (GP) if they were worried about their blood glucose readings, as per usual care. Insulin was adjusted according to need using an insulin adjustment plan, as guided by the CNS. After the 12-week period, the hub was removed from the patients' home (Figure 2).

Figure 2. Intervention. CNS: clinical nurse specialist.



Usual Care

A retrospective audit of 12 patients from 2015 showed that clinic visits ranged from 1 to 3 (mean 1.75) in the 12-month period following the commencement of insulin therapy. Phone calls from the CNS ranged from 0 to 4 per patient (mean 1.7). In this cohort, the mean HbA_{1c} remained relatively stable over 3 collection points—at baseline (T1; 8.5%), 6-month follow-up (8.6%), and 12-month follow-up (8.4%). No clinically or statistically significant changes occurred across the time points.

Data Collection

Data on biomedical variables were collected at a face-to-face meeting with the CNS. HbA_{1c} and BMI data were collected at T1, 12 weeks (at the end of the intervention; T2), and 6 months postbaseline (3 months postintervention; T3). The CNS kept an ongoing telephone call log of insulin dose, patient-reported hypoglycemia symptoms, call frequency and length, and main topics discussed in the calls. In addition, self-efficacy was measured using the Diabetes Empowerment Scale-Short Form [23], and the Diabetes Distress Scale was used to assess diabetes-related emotional distress [24], at the same time points. Satisfaction with the telemonitoring intervention was measured using the Telemedicine Satisfaction and Usefulness Questionnaire at T2 [25]. All T2 and T3 questionnaires were completed remotely via telephone by a researcher not involved in the intervention. All data were entered into the SPSS statistics version 24 (IBM Corp) and checked for accuracy.

All telemonitoring patients were invited to a focus group interview at T2. Two focus group interviews were conducted with 4 and 12 patients in the first and second groups, respectively, at a community health care venue. The aim of each interview was to explore the expectations, achievements, and opinions of the patients of the telemonitoring intervention, and the interviews were facilitated by an experienced interviewer not involved in the intervention, using a topic guide. Once all patients had completed the intervention, the 2 diabetes nurse specialists who coordinated the intervention were interviewed.

Data Analysis

Quantitative Data

Tests of distribution were carried out on all variables before statistical analysis. Variables with a normal distribution were analyzed using the repeated measures analyses of variance test to compare the pre- and postintervention HbA_{1c}, weight, BMI, diabetes empowerment scale (DES) score, and diabetes distress scale (DDS) score at T1, T2, and T3. A *P* value of less than .05 was considered significant. Variables not normally distributed were analyzed using the Friedman test, and if significant, the Wilcoxon test was used to determine which time points are significantly different from T1.

Qualitative Data

Initially, all interviews were openly coded. The descriptive content analysis was carried out using a framework [26].

Emerging codes were examined and compared for any overlap in meaning, and similar codes were collapsed to form higher-order codes. This constant comparative process continued throughout the data analysis process until the major categories that account for the data were developed. The relationships between categories were explored and made explicit. During the analytic process, detailed memos were recorded, which tracked the emerging understandings and the relationship between the categories identified. The findings are reported textually, supported by relevant quotations from the participants. The computer software package NVivo 10 (QSR International) was used to assist in the organization, management, and retrieval of the qualitative data.

Ethics

Ethical approval was received from the regional hospital Research Ethics Committee before commencement of the study. Mechanisms were put in place to ensure that the rights of the participants and their well-being were given precedence over data collection. All proposed participants were given both written and verbal information about the study. Patients were informed about the study face to face and invited to participate. Willing patients were met by the CNS to obtain written consent to participate. All data were anonymized and password protected and stored in accordance with the Data Protection Act, Ireland [27].

Results

Demographics

A convenience sample of 40 patients commenced the intervention, with 1 participant withdrawing early because of connectivity issues. All remaining 39 patients completed the 12-week telemonitoring and completed data collection for all time points postintervention. Table 1 provides the participants' demographic information.

Table 1. Participant demographics (N=39).

Demographics	Values
Age (years)	
Mean	62.4
Range	37-80
Gender, n (%)	
Male	23 (59)
Female	17 (43)

Biomedical Variables, Hemoglobin A_{1c}, Weight, and Insulin Dose

HbA_{1c} (mmol/mol) decreased significantly between T1 and T2 (mean difference [MD] −17.13; P<.001) and T1 and T3 (MD −18.16; P<.001). This represents a 21.2% drop in HbA_{1c} at T3 (Figures 3 and 4). No significant change occurred in weight or BMI over the 6-month period (Table 2). The mean insulin dose at T1 was 17.25 IU (SD 11.1) and ranged from 6 to 60 IU. We audited HbA_{1c} at follow-up in clinic visits in 2019 and found that the changes in HbA_{1c} (mean 60.4 mmol; P<.001) from T1 were maintained.

Figure 3. Changes in mean hemoglobin A_{1c} (mmol/L).

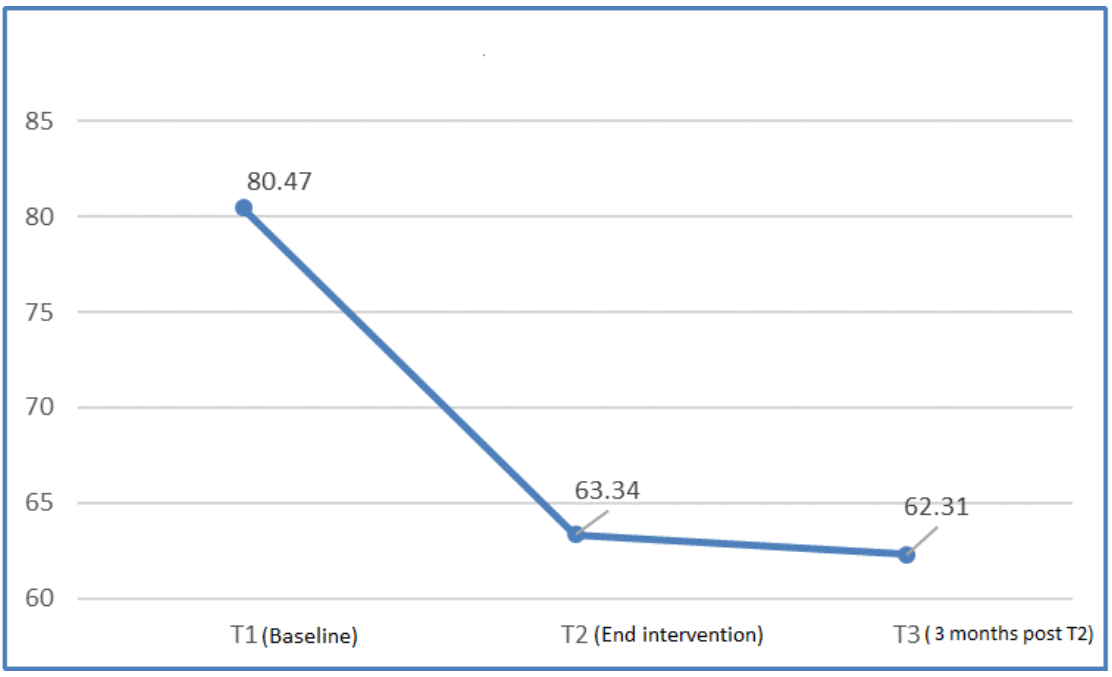


Figure 4. Changes in mean hemoglobin A_{1c} (%).

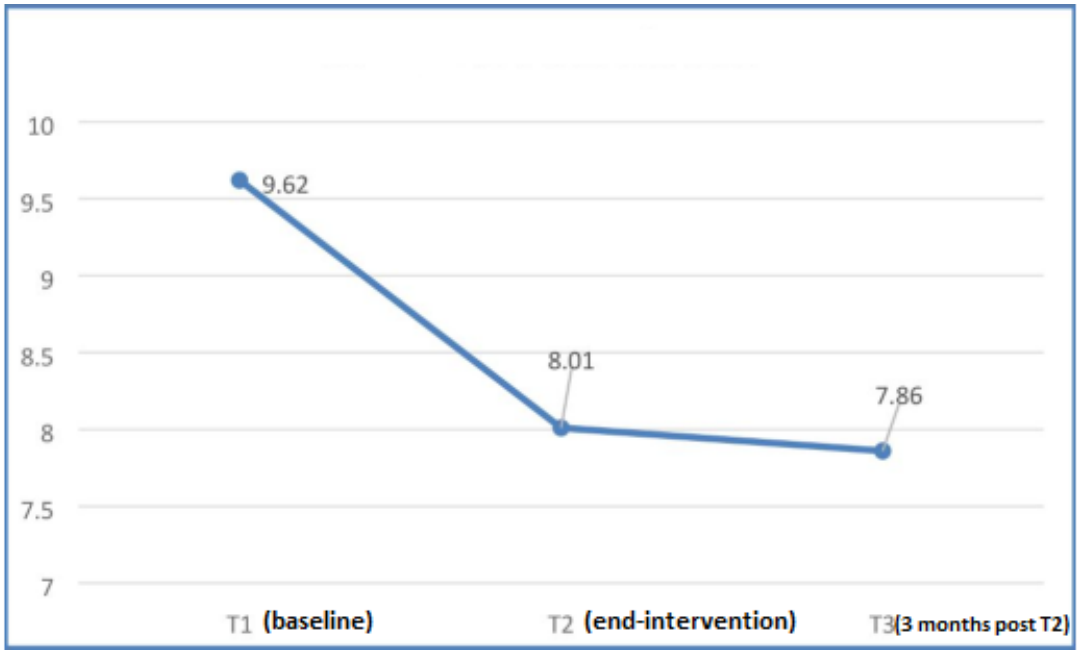


Table 2. Change in weight and BMI among participants (N=39).

Timepoint	Weight (kg), mean (SD)	BMI (kg/m^2), mean (SD)
T1 (baseline)	85.21 (22.75)	30.16 (7.32)
T2 (end of the intervention)	85.18 (20.93)	30.15 (6.82)
T3 (3 months postintervention)	85.63 (21.55)	30.36 (6.97)

Diabetes Empowerment Scale Scores and Diabetes Distress Scale Scores

From T1 to T2, an increase in mean DES score (MD 0.62; $P<.001$) and a decrease in mean DDS score (MD −0.64; $P=.002$) occurred. These differences from T1 were maintained at follow-up at T3 (DES score: MD 0.72; $P<.001$ and DDS score: MD −0.6; $P=.002$; Figures 5 and 6).

Figure 5. Changes in mean total diabetes empowerment scale (DES) scores.

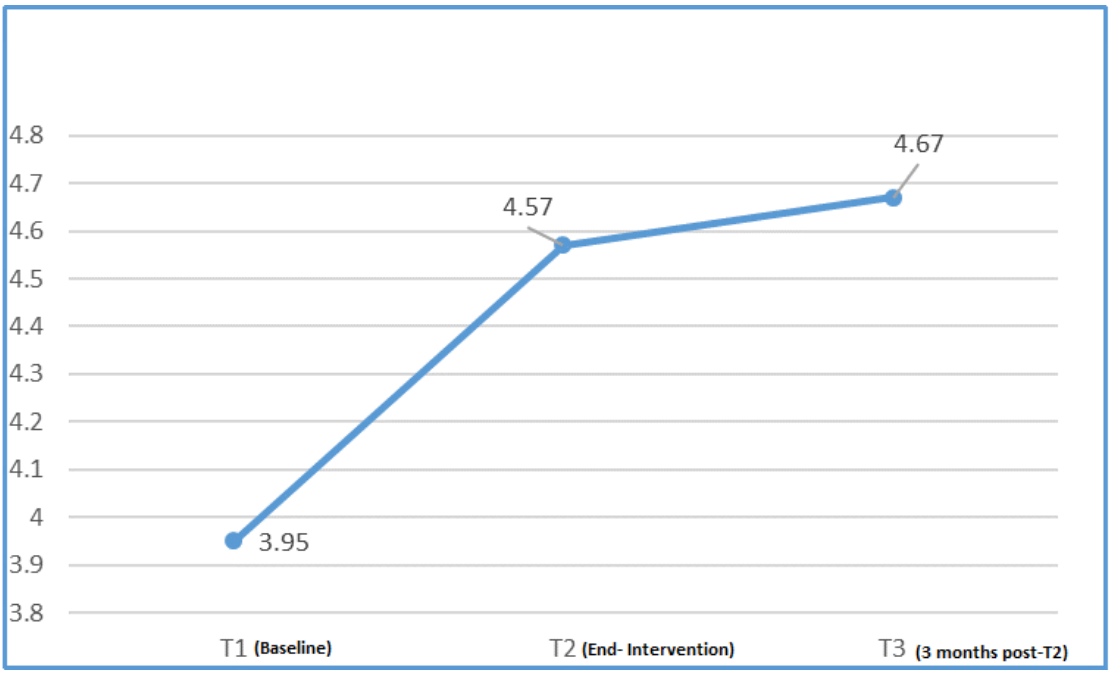
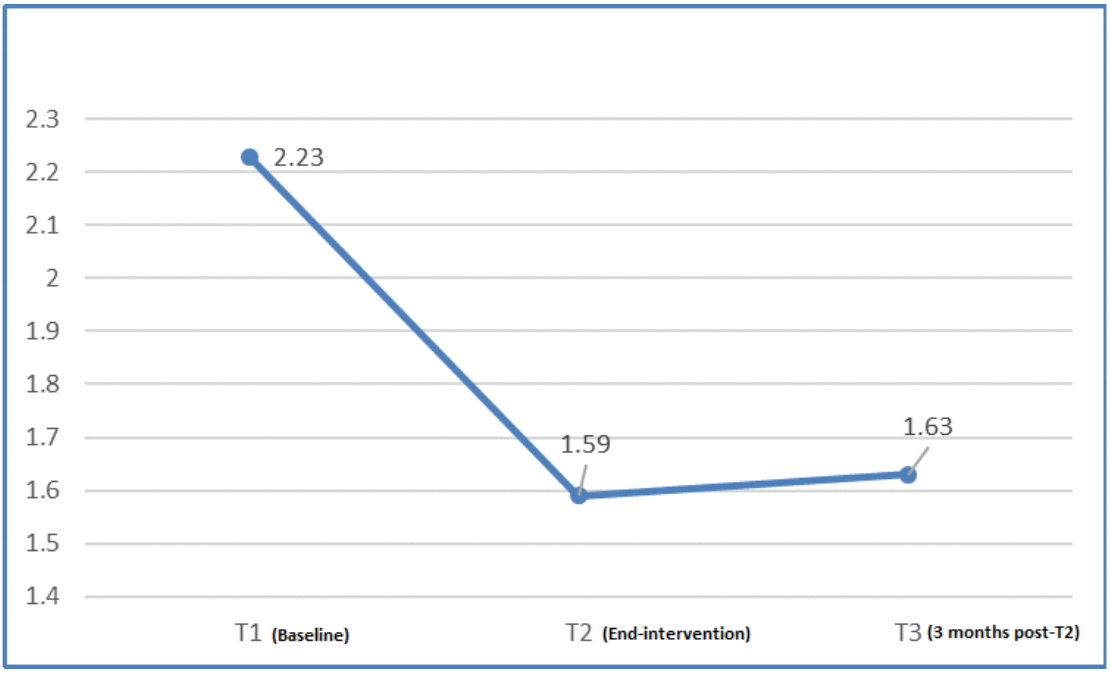


Figure 6. Changes in mean total diabetes distress scale (DDS) scores.



Hypoglycemia

A total of 13% (5/39) of patients had low blood glucose levels during the intervention. Specifically, 3 patients had 2 to 3 episodes, which stopped with advice or dose adjustment. The other 2 patients had multiple episodes, resulting in one patient discontinuing insulin and the other having insulin doses adjusted several times during the study and after. However, during the telephone calls with the CNS, 38% (15/39) of patients reported symptoms of hypoglycemia, demonstrating the need for support for people commencing insulin therapy.

Patient Satisfaction

The mean satisfaction score was above 4 out of a possible 5 on all items on the Telemedicine Satisfaction and Usefulness Questionnaire, indicating a high level of participant satisfaction with the telemonitoring intervention.

Timing and Frequency of Telephone Support

The number and duration of calls were tailored to patients' need. The mean number of calls logged during and after the intervention per participant was 11.5 (SD 4.16). The mean length of the call per participant ranged from 4 min to 15 min. The majority (369/450, 82.0%) of calls were initiated by the CNS, 11.6% (52/450) of calls were initiated by the patients, and 6.2% (28/450) initiation was not documented. The calls reduced dramatically on completion of the intervention, accounting for 8.0% (36/450) after 12 weeks.

The most frequent issue discussed in the telephone calls was poor glucose control (n=176) experienced by the majority of participants (36/39, 92%), followed by injection technique (n=76) reported by 92% (36/39) of participants, patient-reported symptoms of hypoglycemia reported by 38% (15/39) of participants, and self-monitoring of blood glucose technique (n=43) reported by 74% (29/39) of participants. Insulin dosage adjustment (n=147) was the most frequent intervention given to participants (n=37) on one or more occasions by the CNS, followed by healthy eating advice (n=109) and physical activity advice (n=62).

Hospital, General Practitioner, and Diabetes Clinic Visits

Of the total participants, 25% (10/39) attended the diabetic clinic and/or visited the GP and/or experienced an unplanned hospital visit. Of those, 18% (7/39) of participants attended the diabetes clinic, with 2 participants attending twice, and 10% (4/39) of participants visited the GP regarding their diabetes, with 1 participant visiting twice. Unplanned hospital admission was experienced by a minority (3/39, 7%) of the overall number of participants. In the retrospective audit of 12 patients who commenced insulin in 2015, there were 21 clinic visits—we do not have data for this groups' GP visits or hospital admissions.

Qualitative Findings

The thematic analysis of the patient focus group interviews led to 4 main themes: psychological impact of diabetes, increased diabetes empowerment, nurse in the corner, and using the technology. These and their associated subthemes are summarized in Multimedia Appendix 1. The CNS focus group resulted in 5 themes and subthemes, which are summarized in Multimedia Appendix 2: usual transition to insulin, safe transition to insulin therapy using telemonitoring, increased patient empowerment, administrative supports and requirements, and technology. For the focus of this paper, we have selected the themes that relate to patient empowerment for discussion and triangulated the findings from the patient and CNS focus groups and the questionnaire data.

Psychological Impact of Diabetes

In the focus groups, patients talked about the impact of diabetes on their psychological well-being and the need for more support for dealing with the emotions and stress experienced because of the diagnosis and the effects of the illness. They discussed the negative emotions experienced by them because of diabetes,

including anger and agitation with the diagnosis and frustration with the complexity of managing diabetes:

> *There's a lot of agitation and stress around it. For me, in the beginning...when I was diagnosed with diabetes I was very angry because I neither drank nor smoked and I always kept active and working.*

Diabetes Empowerment

In the focus groups, the patients discussed increased self-awareness around the need to self-manage the chronic disease along with developing the knowledge, skills, and confidence to do so. Taking part in the telemonitoring project changed patients' thinking around taking responsibility to manage their diabetes. Having the machine in their home and being responsible for monitoring the blood glucose shifted their thinking:

> *It changes your life forever. It's up to yourself then—do you want it or not. You grab it with both hands.*

Being accountable to someone was a motivating factor in increasing awareness and led them to consider and think about the results they were seeing:

> *It's made me more aware anyway because I'm thinking all the time maybe these bloods are too high, and the nurse is going to say this or maybe they're too low.*

Patients' distress decreased and confidence increased with the comfort of knowing that someone was keeping an eye on the blood glucose all the time:

> *It's giving you control back. You're getting some control over your diabetes when that thing is in the house. Like you say, you can go and download. If you don't feel right about anything in the week that you have done it or in the couple of days, you can go and download it and the nurse will phone you back and reassure you whether it's right or wrong or what you want to do. That's what I like about it.*

Another major source of self-efficacy was mastery experience through taking steps to control diabetes and seeing positive results. In this study, achieving blood glucose control increased their perseverance with managing diabetes:

> *You stick at it. If you're getting good results all the time, you stick at it.*

The patient focus group findings were corroborated by the data from the DES score questionnaires. A significant increase in mean DES score occurred between T1 and T2 (MD 0.62; P=.001) and T1 and T3 (MD 0.72; P<.001; Figure 4).

The CNS team focus group findings also support improved patient self-management with increased patient knowledge and confidence:

> *They're now very confident in managing their insulin so it doesn't stress them if the doctor says we'll just increase that by 2 units. They usually know, and they'll come out and they say to me, I knew we needed to increase that. So, they have gained great confidence in insulin which doesn't happen to everybody.*

The CNS reported problems in the past with the timely uptitration of insulin. Many barriers exist in the usual care of patients to achieving an optimal insulin dose because of not having access to complete and accurate patient data, large caseloads, and insufficient time for the practitioners to frequently contact the patients to adjust their insulin doses. Moreover, 14 patients effectively used an insulin self-adjustment tool, which resulted in earlier titration of insulin to gain timely and improved blood glucose control:

> *I just saw one of them yesterday and she said to me, I did put the insulin dose down a little bit a couple of months ago but actually I had to put it back up again, I didn't ring you because I knew that would be fine...this is a 78-year-old lady. When I say she didn't want to go on insulin, she was absolutely definitely against insulin. To just hear her talking so confidently about her insulin dose and adjusting it and the rationale for the changes she made was great actually.*

Not everyone in the study was taught to self-manage insulin, as it was different from traditional practice and would require further exploration and development for implementation to the wider diabetes population. However, those patients who did use the tool demonstrated increased knowledge and a desire to titrate their own insulin in response to higher blood glucose levels:

> *If it went back up again it would be a case of going to the diabetic nurse and readjusting your insulin. I wouldn't mind if I had to adjust the insulin from now on because you know how it works now.*

Nurse in the Corner

Patients described the telemonitoring as "It's like a nurse in the corner." Having access to the diabetes nurses and knowing that they were reviewing their blood glucose levels gave the patients comfort and a sense of safeness. It also meant that there was quicker intervention when their blood glucose levels were not right, and in their opinion, it reduced their need to visit the hospital and GP.

The added comfort or security relates to the vulnerability the patients felt while commencing insulin. This was an unknown territory, and having close monitoring with professional oversight meant that they felt safe. This links to the development of confidence to manage their diabetes, as an improved emotional state underpins the development of self-efficacy or confidence to manage diabetes:

> *Having it there, I know it's there and I know if I have a problem, I'm going to get the call and the nurse is going to talk me through it. It's hard to explain the comfort that you have in that.*

An Empowering Nurse-Patient Relationship

In the focus group, the CNS participants identified the need for a more equal relationship and enhanced partnership between the nurse and the patient and recognized the role of the

intervention in empowering patients toward self-management. This eased the nurse into the watchful observer role:

> *I suppose that we are always talking about self-management and for the self-management to work it has to be teamwork. It can't be a them and us. You're always trying to build up relationships.*

The CNS perceived that the change in the nurse-patient relationship was toward one of enhanced partnership—a mutual goal that allowed the patient and the nurse to have an equal footing in its achievement:

> *even after the telemonitoring period, we still get phone calls from the patients but it's interesting that there's more an equal relationship with us. It isn't the kind of traditional nurse and patient relationship.*

Increased Workload

CNS participants did highlight that although the telemonitoring system potentially led to much greater efficiencies in terms of patient monitoring and treatment, they strongly voiced the need to have adequate resources to support such a system. One of the key areas of concern was related to adequate nursing resources. As this new telemonitoring system generated large dataset for each patient and these data were being uploaded daily, there was an expectation that nurses would be seeing and reviewing these data regularly. However, because of workload pressures and staff resources, CNS participants reported that often patient data may not be reviewed for several days after upload, and they identified this as a source of concern for patient safety. They would recommend using the track, trend, and triage service, which was not used in this study:

> *the 'buts' are that it definitely added to our work time and I suppose that just the pressure of knowing that you had a responsibility to those patients to look at those readings no matter what else was going on in the service was an additional pressure, I think you would have to restrict the numbers, especially with just the 2 of us; you would have to decide on a certain number at each time. You couldn't just do it all.*

Discussion

Principal Findings

In summary, we observed a significant reduction in HbA_{1c} levels without a significant increase in weight. Patient empowerment scores increased and DDS scores reduced, and these findings were corroborated by the participants' experiences explored in the focus group interviews. These effects were maintained at follow-up 3 months after the intervention ended. Other effects reported by the participants included increased knowledge and competence to self-manage their condition. Some participants engaged in self-titration of insulin using a tool developed by the CNS. A total of 5 patients had episodes of hypoglycemia. A significant increase in workload was reported by the CNS team, which led to several recommendations for streamlining the delivery of telemonitoring with the current service and for additional supports to the health care team.

Patient Empowerment

Patient empowerment in diabetes is fundamental to achieving behavior change, and it is important that the motivation to change is driven internally rather than externally [8]. Self-determination theory proposes that addressing the 3 psychological needs—autonomy, competence, and relatedness—fosters a motivation to engage in healthy behaviors. The extent to which these 3 needs are met or unmet in the social context predicts well-being and thriving [28]. The findings of this study suggest that these 3 psychological needs were met. The CNS observed a positive impact on the nurse-patient relationship, with a balancing out of power and a sense of letting go from the CNS perspective. The perception of increased partnership was empowering to both parties. Patients' autonomy increased with an improved awareness and a heightened level of responsibility for the self-management of diabetes. In a large prospective observational study involving 4341 multinational patients, the quality of the patient-provider relationship was significantly correlated with insulin adherence and HbA_{1c} levels [29].

The patients' knowledge and competence increased in both the effect of their lifestyle behaviors on blood glucose control and how to change them. They attributed this change to the increase in the monitoring of their blood glucose level and the support from the CNS. In a review of qualitative data from seven trials and observational studies of telemonitoring for long-term conditions in primary care, Hanley et al [30] concluded that generating and recording the telemonitoring data had an empowering effect on patients to self-manage. Being active participants in their care increased confidence and enabled them to negotiate prompt access to care.

Insulin Self-Management

Although the central aim of this study was to explore the experience of using telemonitoring to facilitate the transition to insulin therapy, some patients expressed confidence in their knowledge and skill to self-manage their insulin dosage. The CNS participants also indicated an increase in their confidence to let patients self-manage. Empowerment occurs when the goal of the health care practitioner is to enable the patient to critically think and make informed decisions about their care [8]. Self-titration of insulin is well established in type 1 diabetes, but for most patients with type 2 diabetes, dose titration is still carried out by physicians and diabetes nurse specialists. The evidence suggests that this is not the best process to achieve the optimal glycemic control, and self-titration of insulin in type 2 diabetes may be more effective [2]. Computer-assisted self-titration has been found to improve patient awareness of blood glucose management and increase self-efficacy to manage insulin [31]. Our findings suggest that the use of telemonitoring combined with self-titration of insulin empowers people to self-manage and warrants further investigation.

Less Distress, Increased Empowerment Scores, and Fewer Attendances

High levels of diabetes distress negatively affect insulin adherence and glycemic control [29]. The DES scores and DDS scores demonstrated that patient confidence in the ability to

manage diabetes increased and diabetes distress decreased significantly by the end of the 12-week intervention. These findings were supported by the results of the focus group interviews where patients and the CNS team spoke at length about reduced distress and increased patient confidence in their ability to self-manage diabetes. They also reported fewer attendances to the GP and hospital clinics, and we observed a reduction in the number of clinic visits for the intervention group compared with a retrospective cohort. However, the intervention group received a much higher number of calls from the CNS than the retrospective cohort. Telemonitoring and support for people with diabetes have previously demonstrated improvements in self-management and reductions in psychological distress [32].

Glycemic Control

There was a clinically and statistically significant change in HbA_{1c} at 3 months after the end of the telemonitoring intervention to support the introduction of insulin in glucose management, however, with no significant impact on weight. HbA_{1c} reduced by 1.61% at T2 and 1.76% at 6 months from T1. It could be argued that this reduction would usually be observed with the introduction of insulin, and without a control group, it is difficult to posit that the effects are because of the telemonitoring intervention. In a longitudinal study that evaluated the change in HbA_{1c} values after the usual approach to start insulin therapy in 779 patients in primary care practices in Germany and 646 patients in the United Kingdom, with a mean HbA_{1c} of 8.1% (SD 1.3%) and 9.3% (SD 1.5%), respectively [33], the average-adjusted HbA_{1c} improvements in the first 12 months were 0.5% (95% CI 0.4%-0.6%) in Germany and 1.0% (95% CI 0.7%-1.3%) in the United Kingdom. Between 12 and 36 months, these improvements in glycemic control were maintained in both patient groups, without additional improvement in glycemic control. This would suggest that using telemonitoring to initiate insulin is more effective, and in our group, the reduction in mean HbA_{1c} was maintained when we audited the group data in 2019. A randomized controlled trial (RCT) of telemedicine with lifestyle adjustment and titration of metformin plus other oral hypoglycemics or insulin in a Danish type 2 diabetes population demonstrated a 14.6% drop in HbA_{1c} (−15 mmol) with telemedicine and 10.6% (−10 mmol) reduction in their control group [34]. In our retrospective audit of 12 patients commencing insulin in 2015, we saw no clinically significant change in HbA_{1c}, which may be because of the multiple barriers to optimal insulin treatment in primary care at the patient and practitioner levels, and this demonstrates the need for structured self-management support and integrated support systems [4].

Adverse Effects

Some patients did experience hypoglycemia both during and after the intervention, indicating the need for monitoring and support for this population. In an RCT aimed at reducing cardiovascular events in patients with type 2 diabetes using intensive therapy vs standards therapy to reduce HbA_{1c}, the rate of hypoglycemic episodes requiring medical assistance was 3.1% in the intensive therapy group and 1.0% in the standard therapy group, and the mean weight gain at 3 years was 3.5 kg and 0.4 kg in the 2 groups, respectively [35]. In our study, no significant change occurred in weight over the 6-month period. Hypoglycemia and weight gain are among the many recognized worries experienced by both patients and health care practitioners when considering insulin treatment [36]. Although 13% (5/39) of patients had episodes of hypoglycemia, both the patients and the CNS participants felt that telemonitoring increased efficacy and patient safety in the transition to insulin therapy. The intervention allowed more timely adjustment of insulin levels, which leads to faster control of blood sugar levels. Patients reported a reduced need to visit the GP and hospital clinics probably because of the increased level of contact with the CNS during the intervention.

Acceptability

In our study, the use of telemonitoring while transitioning to insulin therapy was highly acceptable to patients, and the use of the technology caused them few problems. There were minor issues with uploading results at the start of the intervention, which could be overcome with an additional visit or better scheduling of the visit from the technology support person.

This intervention has the potential to allow the CNS team to provide more efficient and safe care to the patients, a finding corroborated by Jalil et al [37], but it comes at a cost. The telemonitoring system generated a large dataset for each patient, which had to be reviewed. On reflection, the team felt that they should have availed of the full track, trend, and triage service that is offered as part of the telemonitoring system. This has been found in other studies with one exasperated GP referring to the management of the *tsunami* of patient monitoring data generated by the Whole System Demonstrator project of telemonitoring in long-term conditions in the United Kingdom [38].

The increase in the workload of the team could have been alleviated by administrative support. This would enhance the service by freeing the CNS team to focus on their health care role. There were some teething issues with the introduction of the technology, which caused frustration in both patients and the CNS team. The learning that occurred during the project would allow for a more streamlined approach to the introduction of telemonitoring, with several recommendations for a change to integrating the system with the current service. Although our CNS sample size was small, the findings mirror those of larger studies that have suggested that if the telemonitoring system is not streamlined with the current models of practice, this will impede their uptake by health care professionals [30,38].

Conclusions

The use of telemonitoring while transitioning to insulin therapy was highly acceptable to patients with high satisfaction and increases in confidence and knowledge. The intervention transformed the nurse-patient relationship and resulted in the empowerment of patients to self-manage. Patients reported an increased awareness and level of responsibility and confidence for self-management of diabetes. There was a significant decrease in the DDS score, and both patients and the CNS team reported reduced patient distress. Some patients also expressed

confidence in their knowledge and skill to self-manage their insulin dosage. The CNS team also indicated an increase in their confidence to facilitate patients to self-manage. There was a positive impact on the nurse-patient relationship, with a balancing out of power and a sense of letting go from the CNS team. The increased sense of partnership was empowering to both parties.

There was a clinically and statistically significant drop in HbA_{1c}, as expected with the introduction of insulin in glucose management, with no significant impact on weight and improved sense of patient safety from the patient and practitioner perspective.

Overall, the use of the technology caused a few minor problems for patients. The intervention allowed the CNS team to provide more efficient and safe care to the patients, but it came at a cost. There was a significant increase in the administrative workload of the team, which could have been alleviated by administrative support. There were some teething issues with the introduction of the technology, which caused some frustration in both patients and the CNS team, which may be alleviated with a more streamlined approach and integrating the system with the current service.

Limitations and Further Research

Some limitations of this study need to be considered, and where appropriate, recommendations for further research are proposed. Data in this study were generated through convenience sampling of patients attending one health care region in Ireland, and the number of patients recruited was restricted to 40 by limited nursing and technological resources. In addition, as this was a volunteer sample, these participants may be more motivated toward successful self-management of their diabetes and transition to insulin. Considering this, the findings of this study cannot be generalized to other settings.

The small sample size of both the patient and the CNS groups results in very low study power, which reduces reliability and generalizability. This study design did not allow comparison of this intervention with a control group, which limits the ability to fully determine the effectiveness of the intervention. We audited a retrospective cohort of patients to explore the impact of the intervention; however, we recognize the methodological limitations of this approach. A larger RCT would be useful to allow causal comparisons to be made between telemonitoring-supported care of type 2 diabetes during transition to insulin and conventional care. Measurement of other performance metrics such as frequency of GP visits, levels of technical support needed, and time spent by CNSs in telephone support would also determine the feasibility of such an approach. However, the mixed methods approach used in this study, drawing on both quantitative and qualitative data to support findings, does demonstrate congruence toward patient empowerment experiences when using telemonitoring with telephone support to self-manage diabetes.

Future research design in this area might also consider examining whether improvements in patient outcomes are associated with the effects of using the monitor alone or in combination with increased telephone support and/or the use of an insulin titration scale. Moreover, although this study did look at T1 data and examined the effect at the end of the 12-week intervention and at 3-month follow-up and found positive effects in relation to decreased DDS scores and increased DES scores, further research that would examine if these positive effects were sustained over a longer period would lend strength to the value of such interventions.

Acknowledgments

This study was funded by HSE (West) Ireland Long-Term conditions project.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Participant experiences themes and subthemes.
[DOCX File , 22 KB - jmir_v22i5e16161_app1.docx]

Multimedia Appendix 2
Clinical nurse specialist experiences themes and subthemes.
[DOCX File , 24 KB - jmir_v22i5e16161_app2.docx]

References

1. International Diabetes Federation.: International Diabetes Federation; 2017. IDF Clinical Practice Recommendations for Managing Type 2 Diabetes in Primary Care URL: https://www.idf.org/e-library/guidelines/128-idf-clinical-practice-recommendations-for-managing-type-2-diabetes-in-primary-care.html [accessed 2019-07-10]
2. Khunti K, Davies MJ, Kalra S. Self-titration of insulin in the management of people with type 2 diabetes: a practical solution to improve management in primary care. Diabetes Obes Metab 2013 Aug;15(8):690-700. [doi: 10.1111/dom.12053] [Medline: 23253563]

3. Khunti K, Nikolajsen A, Thorsted BL, Andersen M, Davies MJ, Paul SK. Clinical inertia with regard to intensifying therapy in people with type 2 diabetes treated with basal insulin. Diabetes Obes Metab 2016 Apr;18(4):401-409 [FREE Full text] [doi: 10.1111/dom.12626] [Medline: 26743666]
4. Ellis K, Mulnier H, Forbes A. Perceptions of insulin use in type 2 diabetes in primary care: a thematic synthesis. BMC Fam Pract 2018 May 22;19(1):70 [FREE Full text] [doi: 10.1186/s12875-018-0753-2] [Medline: 29788908]
5. Ng CJ, Lai PS, Lee YK, Azmi SA, Teo CH. Barriers and facilitators to starting insulin in patients with type 2 diabetes: a systematic review. Int J Clin Pract 2015 Oct;69(10):1050-1070. [doi: 10.1111/ijcp.12691] [Medline: 26147376]
6. Khunti K, Damci T, Meneghini L, Pan CY, Yale JF, SOLVE Study Group. Study of Once Daily Levemir (SOLVE): insights into the timing of insulin initiation in people with poorly controlled type 2 diabetes in routine clinical practice. Diabetes Obes Metab 2012 Jul;14(7):654-661. [doi: 10.1111/j.1463-1326.2012.01602.x] [Medline: 22443213]
7. Hsu WC, Lau KH, Huang R, Ghiloni S, Le H, Gilroy S, et al. Utilization of a cloud-based diabetes management program for insulin initiation and titration enables collaborative decision making between healthcare providers and patients. Diabetes Technol Ther 2016 Feb;18(2):59-67 [FREE Full text] [doi: 10.1089/dia.2015.0160] [Medline: 26645932]
8. Anderson RM, Funnell MM. Patient empowerment: myths and misconceptions. Patient Educ Couns 2010 Jun;79(3):277-282 [FREE Full text] [doi: 10.1016/j.pec.2009.07.025] [Medline: 19682830]
9. Chen SY, Hsu HC, Wang RH, Lee YJ, Hsieh CH. Glycemic control in insulin-treated patients with type 2 diabetes: empowerment perceptions and diabetes distress as important determinants. Biol Res Nurs 2019 Mar;21(2):182-189. [doi: 10.1177/1099800418820170] [Medline: 30585081]
10. Long AF, Gambling T. Enhancing health literacy and behavioural change within a tele-care education and support intervention for people with type 2 diabetes. Health Expect 2012 Sep;15(3):267-282 [FREE Full text] [doi: 10.1111/j.1369-7625.2011.00678.x] [Medline: 21521430]
11. Chen L, Chuang L, Chang C, Wang C, Wang I, Chung Y, et al. Evaluating self-management behaviors of diabetic patients in a telehealthcare program: longitudinal study over 18 months. J Med Internet Res 2013 Dec 9;15(12):e266 [FREE Full text] [doi: 10.2196/jmir.2699] [Medline: 24323283]
12. Carlisle K, Warren R. A qualitative case study of telehealth for in-home monitoring to support the management of type 2 diabetes. J Telemed Telecare 2013 Oct;19(7):372-375. [doi: 10.1177/1357633X13506512] [Medline: 24218347]
13. Shane-McWhorter L, Lenert L, Petersen M, Woolsey S, McAdam-Marx C, Coursey JM, et al. The Utah Remote Monitoring Project: improving health care one patient at a time. Diabetes Technol Ther 2014 Oct;16(10):653-660 [FREE Full text] [doi: 10.1089/dia.2014.0045] [Medline: 24991923]
14. Risling T, Martinez J, Young J, Thorp-Froslie N. Evaluating patient empowerment in association with eHealth technology: scoping review. J Med Internet Res 2017 Sep 29;19(9):e329 [FREE Full text] [doi: 10.2196/jmir.7809] [Medline: 28963090]
15. Risling T, Martinez J, Young J, Thorp-Froslie N. Defining empowerment and supporting engagement using patient views from the citizen health information portal: qualitative study. JMIR Med Inform 2018 Sep 10;6(3):e43 [FREE Full text] [doi: 10.2196/medinform.8828] [Medline: 30201603]
16. Riley JP, Gabe JP, Cowie MR. Does telemonitoring in heart failure empower patients for self-care? A qualitative study. J Clin Nurs 2013 Sep;22(17-18):2444-2455. [doi: 10.1111/j.1365-2702.2012.04294.x] [Medline: 23185992]
17. Jaana M, Sherrard H, Paré G. A prospective evaluation of telemonitoring use by seniors with chronic heart failure: Adoption, self-care, and empowerment. Health Informatics J 2019 Dec;25(4):1800-1814. [doi: 10.1177/1460458218799458] [Medline: 30247080]
18. Ammenwerth E, Woess S, Baumgartner C, Fetz B, van der Heidt A, Kastner P, et al. Evaluation of an integrated telemonitoring surveillance system in patients with coronary heart disease. Methods Inf Med 2015;54(5):388-397. [doi: 10.3414/ME15-02-0002] [Medline: 26395147]
19. Shulman KI. Clock-drawing: is it the ideal cognitive screening test? Int J Geriatr Psychiatry 2000 Jun;15(6):548-561. [doi: 10.1002/1099-1166(200006)15:6<548::aid-gps242>3.0.co;2-u] [Medline: 10861923]
20. Health Service Executive Ireland (HSE). Guide to Blood Glucose Testing in Type 2 Diabetes. Ireland: HSE; 2015. URL: http://www.hse.ie/eng/about/Who/clinical/natclinprog/diabetesprogramme/glucosetesting.pdf [accessed 2015-11-15]
21. Turner J, Larsen M, Tarassenko L, Neil A, Farmer A. Implementation of telehealth support for patients with type 2 diabetes using insulin treatment: an exploratory study. Inform Prim Care 2009;17(1):47-53 [FREE Full text] [doi: 10.14236/jhi.v17i1.714] [Medline: 19490773]
22. Davies MJ, D'Alessio DA, Fradkin J, Kernan WN, Mathieu C, Mingrone G, et al. Management of hyperglycaemia in type 2 diabetes, 2018. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetologia 2018 Dec;61(12):2461-2498. [doi: 10.1007/s00125-018-4729-5] [Medline: 30288571]
23. Anderson RM, Fitzgerald JT, Gruppen LD, Funnell MM, Oh MS. The Diabetes Empowerment Scale-Short Form (DES-SF). Diabetes Care 2003 May;26(5):1641-1642. [doi: 10.2337/diacare.26.5.1641-a] [Medline: 12716841]
24. Polonsky WH, Fisher L, Earles J, Dudl RJ, Lees J, Mullan J, et al. Assessing psychosocial distress in diabetes: development of the diabetes distress scale. Diabetes Care 2005 Mar;28(3):626-631. [doi: 10.2337/diacare.28.3.626] [Medline: 15735199]

25. Bakken S, Grullon-Figueroa L, Izquierdo R, Lee N, Morin P, Palmas W, IDEATel Consortium. Development, validation, and use of English and Spanish versions of the telemedicine satisfaction and usefulness questionnaire. J Am Med Inform Assoc 2006;13(6):660-667 [FREE Full text] [doi: 10.1197/jamia.M2146] [Medline: 16929036]
26. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
27. Irish Statute Book. 2018. Data Protection Act 2018 URL: http://www.irishstatutebook.ie/eli/2018/act/7/enacted/en/print#sec1 [accessed 2017-05-05]
28. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
29. Linetzky B, Jiang D, Funnell MM, Curtis BH, Polonsky WH. Exploring the role of the patient-physician relationship on insulin adherence and clinical outcomes in type 2 diabetes: Insights from the MOSAIc study. J Diabetes 2017 Jun;9(6):596-605. [doi: 10.1111/1753-0407.12443] [Medline: 27368146]
30. Hanley J, Pinnock H, Paterson M, McKinstry B. Implementing telemonitoring in primary care: learning from a large qualitative dataset gathered during a series of studies. BMC Fam Pract 2018 Jul 18;19(1):118 [FREE Full text] [doi: 10.1186/s12875-018-0814-6] [Medline: 30021535]
31. Simon AC, Gude WT, Holleman F, Hoekstra JB, Peek N. Diabetes patients' experiences with the implementation of insulin therapy and their perceptions of computer-assisted self-management systems for insulin therapy. J Med Internet Res 2014 Oct 23;16(10):e235 [FREE Full text] [doi: 10.2196/jmir.3198] [Medline: 25340869]
32. Aikens JE, Rosland A, Piette JD. Improvements in illness self-management and psychological distress associated with telemonitoring support for adults with diabetes. Prim Care Diabetes 2015 Apr;9(2):127-134. [doi: 10.1016/j.pcd.2014.06.003] [Medline: 25065270]
33. Rathmann W, Strassburger K, Tamayo T, Kostev K. Longitudinal change in HbA1c after insulin initiation in primary care patients with type 2 diabetes: a database analysis in UK and Germany. Prim Care Diabetes 2012 Apr;6(1):47-52. [doi: 10.1016/j.pcd.2011.10.001] [Medline: 22099215]
34. Rasmussen OW, Lauszus FF, Loekke M. Telemedicine compared with standard care in type 2 diabetes mellitus: A randomized trial in an outpatient clinic. J Telemed Telecare 2016 Sep;22(6):363-368. [doi: 10.1177/1357633X15608984] [Medline: 26468213]
35. Action to Control Cardiovascular Risk in Diabetes Study Group, Gerstein HC, Miller ME, Byington RP, Goff DC, Bigger JT, et al. Effects of intensive glucose lowering in type 2 diabetes. N Engl J Med 2008 Jun 12;358(24):2545-2559 [FREE Full text] [doi: 10.1056/NEJMoa0802743] [Medline: 18539917]
36. Home P, Riddle M, Cefalu WT, Bailey CJ, Bretzel RG, Del Prato S, et al. Insulin therapy in people with type 2 diabetes: opportunities and challenges? Diabetes Care 2014 Jun;37(6):1499-1508 [FREE Full text] [doi: 10.2337/dc13-2743] [Medline: 24855154]
37. Jalil S, Myers T, Atkinson I, Soden M. Complementing a clinical trial with human-computer interaction: patients' user experience with Telehealth. JMIR Hum Factors 2019 Jun 6;6(2):e9481 [FREE Full text] [doi: 10.2196/humanfactors.9481] [Medline: 31172958]
38. MacNeill V, Sanders C, Fitzpatrick R, Hendy J, Barlow J, Knapp M, et al. Experiences of front-line health professionals in the delivery of telehealth: a qualitative study. Br J Gen Pract 2014 Jul;64(624):e401-e407 [FREE Full text] [doi: 10.3399/bjgp14X680485] [Medline: 24982492]

Abbreviations

CNS: clinical nurse specialist
DDS: diabetes distress scale
DES: diabetes empowerment scale
eHealth: electronic health
GP: general practitioner
HbA_{1c}: hemoglobin A_{1c}
HSE: Health Service Executive
MD: mean difference
RCT: randomized controlled trial
T1: baseline
T2: end of the intervention
T3: 3 months postintervention

Edited by G Eysenbach; submitted 06.09.19; peer-reviewed by B McKinstry, K Fitzner; comments to author 18.10.19; revised version received 23.12.19; accepted 24.01.20; published 14.05.20.

Please cite as:
McGloin H, O'Connell D, Glacken M, Mc Sharry P, Healy D, Winters-O'Donnell L, Crerand K, Gavaghan A, Doherty L
Patient Empowerment Using Electronic Telemonitoring With Telephone Support in the Transition to Insulin Therapy in Adults With Type 2 Diabetes: Observational, Pre-Post, Mixed Methods Study
J Med Internet Res 2020;22(5):e16161
URL: https://www.jmir.org/2020/5/e16161
doi:10.2196/16161
PMID:32406854

Original Paper

Clinical and Technical Considerations of an Open Access Telehealth Network in South Carolina: Definition and Deployment

Kathryn King[1], MD, MHS; Dee Ford[2], MD, MSCR; Michael Haschker[3], AE; Jillian Harvey[4], PhD, MPH; Ryan Kruis[3], MSW; James McElligott[3], MD, MSCR

[1]Department of Pediatrics, Medical University of South Carolina, Charleston, SC, United States

[2]Department of Pulmonary and Critical Care Medicine, Medical University of South Carolina, Charleston, SC, United States

[3]Center for Telehealth, Medical University of South Carolina, Charleston, SC, United States

[4]Department of Healthcare Leadership and Management, Medical University of South Carolina, Charleston, SC, United States

Corresponding Author:
Jillian Harvey, PhD, MPH
Department of Healthcare Leadership and Management
Medical University of South Carolina
151B Rutledge Ave
MSC 962
Charleston, SC, 29425
United States
Phone: 1 8437923431
Email: harveyji@musc.edu

Abstract

Background: Today, telehealth is experiencing exponential growth in utilization. Paralleling this trend is the growth in the telehealth industry, with sharp increases in the number of platforms, functionalities, and levels of integrations within both the electronic health record and other technical systems supporting health care. When a telehealth network is intended to be used across independent health care systems, an additional layer of complexity emerges. In the context of regionalized telehealth networks that are not within the same health care system, not only are technical interoperability challenges a practical barrier, but administrative, clinical, and competitive elements also quickly emerge, resulting in fragmented, siloed technologies.

Objective: The study aimed to describe a statewide approach to deploying an interoperable open access telehealth network across multiple health systems.

Methods: One promising solution to the abovementioned concerns is an open access telehealth network. In the field of telehealth, an open access network (OAN) can be defined as a network infrastructure that can be used by health care providers without a closed or proprietary platform, specific obligatory network, or service-specific telehealth technologies. This framework for the development of an OAN is grounded in practical examples of clinical programs that function in each stage of network maturity based on the experience of the South Carolina Telehealth Alliance (SCTA). The SCTA's experience details successes and challenges in an ongoing effort to achieve an OAN. The model describes an OAN in stages of collaborative maturity and provides insights into the technological, clinical, and administrative implications of making the collaboration possible.

Results: The four stages of an OAN are defined according to operational maturity, ranging from feasibility to demonstration of implementation. Each stage is associated with infrastructure and resource requirements and technical and clinical activities. In stage 1, technical standards are agreed upon, and the clinical programs are designed to utilize compliant technologies. In stage 2, collaboration is demonstrated through technical teams working together to address barriers, whereas clinical and administrative teams share best practices. In stage 3, a functional interoperable network is demonstrated with different institutions providing service through common telehealth end points at different patient care sites. In stage 4, clinical workflows are streamlined and standardized across institutions, and economies of scale are achieved through technical and administrative innovations.

Conclusions: The approach to OAN development described provides a roadmap for achieving a functional telehealth network across independent health systems. The South Carolina experience reveals both successes and challenges in achieving this goal. The next steps toward the development of OANs include advocacy and ongoing engagement with the developers of telehealth technologies regarding their commitment to interoperability.

(J Med Internet Res 2020;22(5):e17348) doi:10.2196/17348

KEYWORDS

telehealth; health information technology

Introduction

Background

Today, telehealth is experiencing exponential growth in utilization [1,2]. Paralleling this trend is the growth in the telehealth industry, with sharp increases in the number of platforms, functionalities, and levels of integrations within both the electronic health record (EHR) and other technical systems supporting health care. When a telehealth network is intended to be used across independent health care systems, an additional layer of complexity emerges. In the context of regionalized telehealth networks that are not within the same health care system, not only are technical interoperability challenges a practical barrier, but administrative, clinical, and competitive elements also quickly emerge, resulting in fragmented, siloed technologies. These observations echo those of the rapid deployment of EHRs that led to fragmentation and barriers to interoperability across systems that have become a practical issue and barrier to optimal care for patients, providers, and health systems [3]. As rapid deployment of telehealth technology follows a similar trajectory, it becomes increasingly important to learn from the EHR example and develop roadmaps for telehealth collaborative solutions before siloed technology becomes standard practice.

Open Access Network

One promising solution to these concerns is an open access telehealth network. In the field of telehealth, an open access network (OAN) can be defined as a network infrastructure that can be used by health care providers without a closed or proprietary platform, specific obligatory network, or service-specific telehealth technologies. An OAN functioning at the highest level of maturity would also include clinical and administrative workflow standardization. Many manufacturers offer turnkey solutions for starting telehealth programs by selling closed systems and/or proprietary technologies. Although these solutions address immediate needs, they also create closed or siloed networks that cannot be easily accessed or expanded and lack interoperability with other telehealth solutions.

An OAN benefits a health care system as a whole by mitigating the need for proprietary equipment and the specialized staff and contracts that support such equipment. In addition, individual institutions and broader regional health care systems benefit from an OAN as they are able to connect to a more extensive array of subspecialty providers using the same technology, thus saving space, cost, and time in the deployment of equipment. Finally, both providers and patients benefit from familiarity with standardized equipment, increasing adoption for these pivotal end user groups. Open platforms ensure that additional consideration is given to maintaining reasonable costs for broad participation among regional health systems.

Similar to the experience of EHR development, the need for telehealth interoperability was acknowledged nearly a decade ago [4], though the reality of interoperability and effectiveness are not well described [5]. Although the development of an interoperable telehealth network is less complex than a full health information exchange, there are many parallels to consider. The National Quality Forum introduced measures for health information exchange interoperability (Table 1), and the majority of measures go beyond whether the systems are technically interoperable and instead focus on whether the system is used as intended and if that use is effective [5]. In much the same way, a telehealth network may be built to be interoperable, but health systems may choose to selectively deploy this function. For instance, these systems are likely to make variable investments in their own telehealth efforts that are reflective of their variable business cases for making technical decisions. It should be no surprise then that during the growth of telehealth there may also be the need to focus on metrics that relate to interoperability. In this paper, a statewide approach to deploying an interoperable OAN across multiple health systems is described.

Table 1. National Quality Forum domains and subdomains of interoperability.

Domain	Subdomains
Exchange of EHI[a]	• Availability of EHI • Quality of data • Method of exchange
Usability of exchanged EHI	• Relevance • Accessibility • Comprehensibility
Application of exchanged EHI	• Human use • Computable
Impact of interoperability	• Patient safety • Cost savings • Productivity • Care coordination • Improved health care processes and health outcomes • Patient/caregiver engagement • Patient/caregiver experience

[a]EHI: electronic health information.

Methods

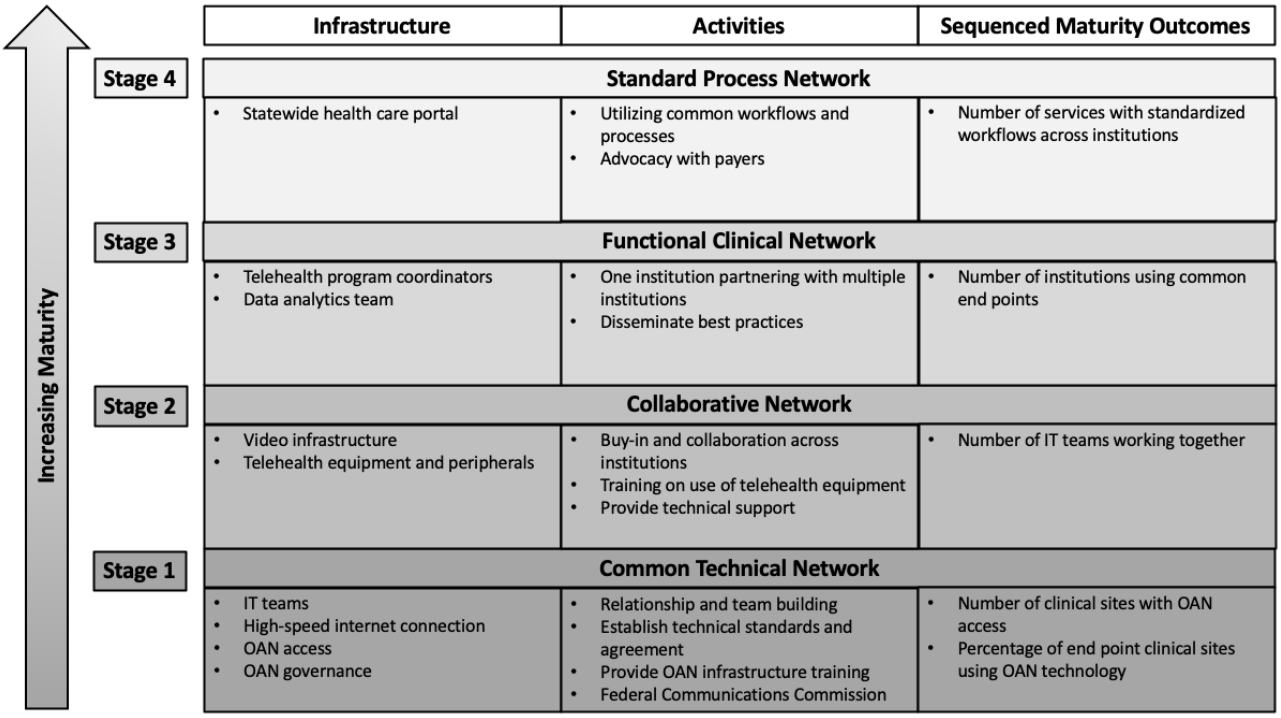
The framework for the development of an OAN is grounded in practical examples of clinical programs that function in each stage of network maturity. Although the model is intended to be generalizable, the experience of the South Carolina Telehealth Alliance (SCTA) is used as a representative use case. The SCTA was established in 2013 as the product of legislatively appropriated state support for telehealth in South Carolina. The multistakeholder alliance includes representatives from local health care systems and payers and has a mission to support the delivery of high-value telehealth across the state. Across South Carolina's 46 counties, there are currently over 400 sites equipped for telehealth services. The SCTA's experience details successes and challenges in an ongoing effort to achieve an OAN. The model presented here describes an OAN in stages of collaborative maturity and provides insights into the technological, clinical, and administrative implications of making the collaboration possible.

Results

Developmental Stages of the Open Access Network

In South Carolina, a vision for technologic interoperability was and remains a key strategy of the state-sponsored telehealth network [6]. The aspirational goal is to allow access to all providers wishing to leverage any deployed telehealth technology in the state. Although achieving this goal is an ongoing challenge, South Carolina has made substantial progress and learned important lessons. A four-stage process is being used to develop and mature a statewide OAN (Figure 1). These four stages are defined according to operational maturity, ranging from feasibility to demonstration of implementation. Each stage is associated with infrastructure and resource requirements, technical, and clinical activities. In stage 1, technical standards are agreed upon, and the clinical programs are designed to utilize compliant technologies. In stage 2, collaboration is demonstrated through technical teams working together to address barriers, whereas clinical and administrative teams share best practices. In stage 3, a functional interoperable network is demonstrated with different institutions providing service through common telehealth end points at different patient care sites. In the fourth and final stage, clinical workflows are streamlined and standardized across institutions, and economies of scale are achieved through technical and administrative innovations (eg, common scheduling portals and standard contract language; Figure 1).

Figure 1. Proposed model for maturing a telehealth open access network. FCC: Federal Communications Commission; IT: information technology; OAN: open access network.



Stage 1: Establish a Common Technical Network

Establish a Collaborative Forum

As in all collaborative efforts that involve multiple stakeholders, establishing a forum for communication, discussion, and gaining trust among the parties is essential. These stakeholders are likely to be asked to accommodate some adjustments internally for the good of the network; thus, it is important to involve them at early stages of strategic planning and the setting of short- and long-term goals. In South Carolina, the SCTA engages stakeholders with an advisory council and a system of workgroups that operate under the alignment of a formal strategic plan [6]. These workgroups, made up of personnel from the various involved institutions, are responsible for operationalizing the plan within their respective organizations and reporting back to the advisory council. The SCTA information technology (IT) workgroup has become a common ground for members to build relationships and make collaborative decisions with the common goal of interoperability, scalability, and support of the network.

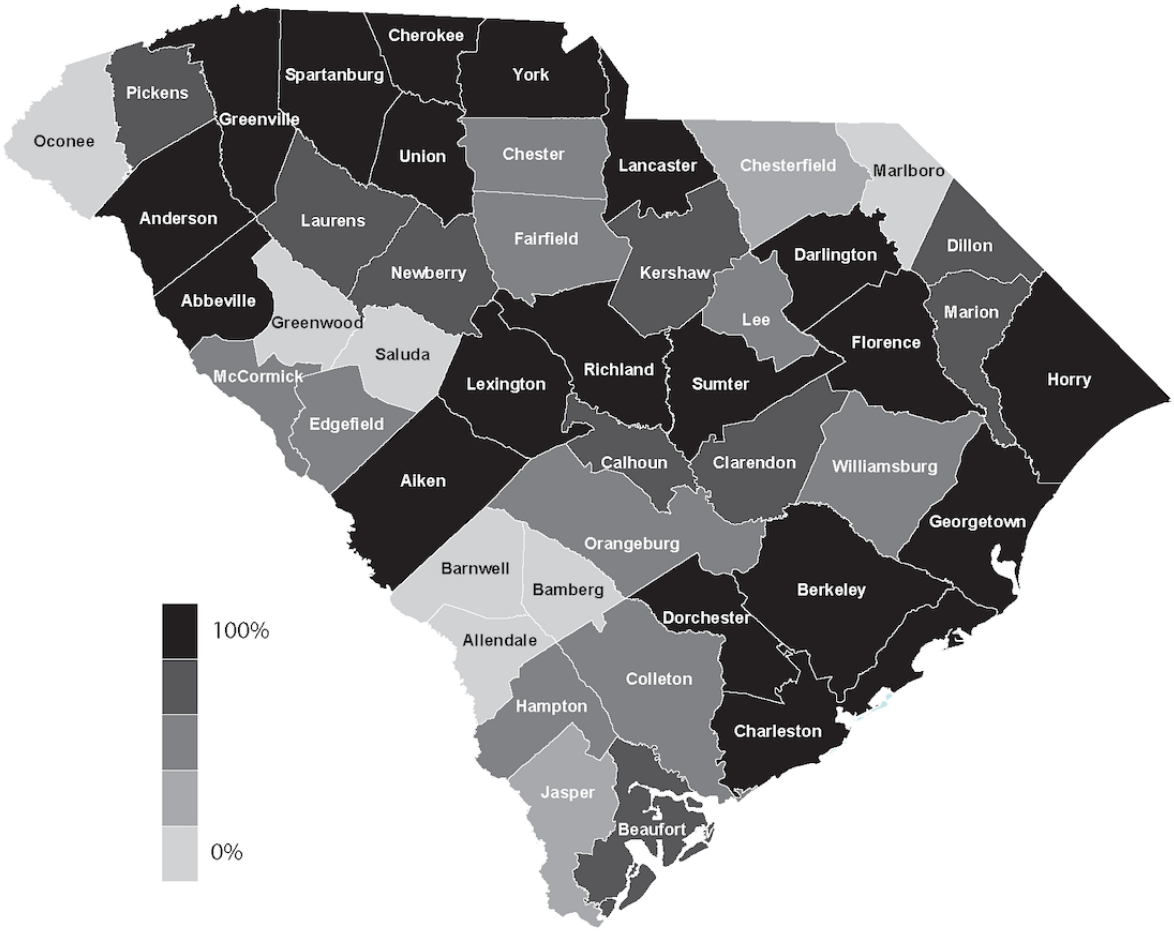
Establish Technical Standards and Agreement

As discussed, the inception of a regional telehealth network requires careful planning and coordination of efforts. Defining network requirements and functionality helps drive important decisions on technical protocols and the types of equipment used. The backbone of the OAN in South Carolina is a video network whose interoperability is made possible by using standardized specifications. These specifications are based upon open standards as defined by the International Telecommunications Union (ITU) and via a collaborative approach defined by the IT workgroup within the SCTA [7]. ITU open standards address specific requirements for functionality, interoperability, and compatibility. Through the utilization of existing infrastructure, the network leveraged investment protection while allowing sites and health care providers choices in the platforms they would use if they met the ITU open systems standards.

Assess and Address Broadband Availability

In South Carolina, rural health care providers often lack technical support, person power, and adequate broadband access. This limited connectivity compounds existing health care access issues, particularly in a region that is mostly rural (Figure 2). Although broadband access is beyond the scope of this paper, it is an important element to address when considering an OAN. When South Carolina designed and implemented a statewide telehealth network, one of the primary requirements was the ability to be used across rural providers and independent health care systems. Many of these health care systems had existing infrastructure and isolated networks that could be reconfigured to work together as one large statewide telehealth network. To build the network, South Carolina utilized the Federal Communications Commission's (FCC) rural health care program that subsidized infrastructure and broadband connectivity, effectively creating a dedicated statewide health care network. The network is known as the Palmetto State Providers Network (PSPN) [8]. Today, PSPN has over 300 participating health care sites providing affordable low-latency dedicated bandwidth with access to secure encrypted video infrastructure for the delivery of telehealth applications and services.

Figure 2. Federal Communications Commission broadband availability in South Carolina.



Open Access Network User End Points

Telehealth end points are hardware platforms that users control to establish connections between referring and consulting telehealth participants. An end point is where physical and logical connections to video and audio systems such as cameras, speakers, and microphones are made. End points are the *end of the line* for the video conferencing connection and are different from other components of a video conferencing network such as gateways or multipoint conference units (bridges), which are video infrastructure components.

End points used in telehealth settings must be dependable and easy to use and meet the needs of the patient and health care provider. Under the OAN system, the patient and provider are not limited to consultations within a specific platform. There are many platforms/end point choices that promote technological efficiencies and streamlined patient care. End points come in a number of platforms and include software and PC-based, tablets, wall-mounted hardware-based units, dedicated desktop units, and the popular clinical cart units widely used in hospital settings. In health care facilities where space and resources are at a premium, the ability to utilize end points across multiple telehealth services and platforms is essential to the growth of the telehealth provision and utilization. The South Carolina approach to end point decisions is to educate providers on what is available, listen to the provider's needs, and then recommend the best solution. Ultimately, it is the health care provider's decision on what is the best end point for the delivery of health care services.

Open Access Network Technical Support

With the deployment of health care networks that provide services for critical and acute care patients, technical reliability is mandatory. Choosing industry-proven commercial-grade network equipment is an essential component for reliability. Having service-level agreements with the manufacturers and providers of network equipment is also a critical element to building and operating a strong network. Knowing that all equipment fails over time, managing equipment life cycles, and providing routine scheduled maintenance and upgrades are essential in maintaining a robust telehealth network. Networks

built with designed redundancy or high availability offer both hardware protection and the ability to perform maintenance and upgrades anytime, day or night. Utilizing site and service agreements, the SCTA offered a collaborative approach to technical support where the regional health care systems provide technical and site support to rural providers and smaller clinics in their region.

Open Access Network and Telehealth Software

One primary question faced when telehealth software is applied within an OAN is whether the software restricts the type of end point that can be utilized. A key component of this question is whether the telehealth end points not manufactured by the vendor can be used when the vendor's software initiates the video call. In addition, there are multiple questions that should be collaboratively discussed and considered before selecting telehealth software (Table 2). Conversely, it is important to consider whether the end points that are manufactured by the vendor can be utilized without the vendor's software. Many national telehealth vendors are moving toward interoperability with computers and tablet-based devices, and many video conferencing products can be configured to call video solutions across a range of video clients. When applying telehealth software, which leverages or augments video conferencing, there are typically two broad uses: encounter management and video network management. Encounter management involves assistance with the workflows of the clinical process; this may include gathering of data to assist clinical decision making, such as radiological imaging to assist in the assessment of a stroke. Encounter assistance may also include administrative functions of care, such as scheduling, facilitating collaborative documentation between care teams, or providing information for billing purposes. However, the functionality of the end points or the software remains constricted by certain specifications determined by telehealth software and equipment manufacturers. For instance, many telehealth encounters are facilitated with the use of far-end camera control in which the distant provider can simulate the in-person experience by controlling the camera on the patient's side. The ability to leverage this functionality may depend on the functionality of telehealth support software or on the specific camera included in the end point. Restrictions in functionality may also be experienced for the use of examination peripherals, which exclude some services from being offered as the appropriate standard of care.

Table 2. Questions to consider when evaluating technologies for users of the open access network.

Category	Questions
Infrastructure	• What problem are we solving? • What are the current video client- or cloud-based services in use? • Does the technology require changes to existing OAN[a] infrastructure? • Does the technology require additional infrastructure?

OAN standards/compatibility	• Will functionality be lost when crossing between software and certain standards-based endpoints in the network? Specifically:
	  • Far-end camera control
	  • Peripheral examinations devices including examination cameras and stethoscopes
	• Are there any end points that software apps provided by a vendor cannot connect to via video?
	• Are the new technologies compatible with the OAN?

Technical support/outcomes	• What are the technical support requirements? • Is there the ability to monitor the end points to ensure they are online when clinically needed? • Is the technology able to provide feedback to monitor key process metrics and data to support relevant quality and outcome measures?
Legal/compliance	• What are the state and federal regulatory requirements (including Health Insurance Portability and Accountability Act, Family Educational Rights and Privacy Act, and Commission on Accreditation for Law Enforcement Agencies)? • What are the information security requirements addressing the use of video technologies in health care settings? • What end point encryption is required?

[a]OAN: open access network.

The process of selecting technology for use within the OAN requires careful planning to ensure interoperability and appropriateness for each use case. Table 2 outlines several categories of questions that may be helpful to consider in this process.

Stage 1 Example: Connecting Medication-Assisted Therapy Providers to Drug and Alcohol Treatment Centers

An illustrative example of the South Carolina program that is in stage 1 of OAN maturity is that of the telehealth-enabled medication-assisted therapy program that connects eligible prescribers to local drug and alcohol treatment centers to treat patients with opioid use disorder. As the need for this clinical service was identified, the key stakeholders met to establish a collaborative forum to discuss the questions listed in Table 2 and apply the OAN standards that had been established by the SCTA technology workgroup. Broadband capability was assessed at each of the drug and alcohol treatment centers that were provided with federally subsidized broadband, when necessary. The program utilizes standards-based video conferencing equipment and software to connect the patient with contracted providers, and technical support is provided by

a common SCTA support partner. As the program matures, workgroups are forming to progress toward clinical and administrative consensus.

Stage 2: Collaborative Network

Although agreeing upon technical standards is an essential first step, the challenging process of collaboration must be a continuous area of focus to achieve the purposes of the OAN.

Ongoing Technical Collaboration

The implementation of a regional telehealth network requires a high degree of collaboration among the participating sites and service providers who may be unaccustomed to external collaborations. The coordination of efforts on a shared vision for a telehealth network is crucial for successful implementation and deployment. Regulatory requirements and threats to IT security create an environment where IT staff are very protective of their organization's networks. These factors create challenges to establishing trust and working together for the common goal of a statewide health care network. In South Carolina, a three-step process was used to develop and mature statewide IT relationships. In step 1 through the SCTA, we formed an inclusive statewide IT workgroup with representation from members of participating sites. Representation included participation from rural health care providers, community hospitals, regional hospitals, and the state's largest medical centers. In step 2, the workgroup agreed upon a common vision for the network and developed standards used to implement the network with a focus on protecting existing infrastructure and end point investments. In step 3, key stakeholders worked together to enhance the network by addressing needs for accessing technical support, online directory services for end points, and a shared on-call pool for emergent support needs. Working collaboratively and building trust are the cornerstones for developing and implementing a statewide OAN.

Collaborative Design in the Context of Performance and Security

Network Security

With federal and state mandates for the protection of health care information, South Carolina designed the OAN to utilize encryption technologies for all video calls. End points across multiple institutions and clinical sites are configured to standards that include required encryption configurations. Laptops utilized in the delivery of telehealth have encryption technologies installed to protect the laptop and its data from unauthorized access. User accounts are needed for access to all PC-based video clients.

Network Management

Successful design, implementation, and support of a telehealth network is not complete without a comprehensive network management policy and process. Network management should address procedures for the following:

1. Service level agreements on uptime, trouble resolution, and customer expectations.
2. Asset and inventory management
3. Proactive real-time monitoring of all critical systems
4. Equipment life cycle management
5. Problem escalation procedures
6. Equipment manufacture support agreements
7. Ticketing system for reporting, resolving, and managing trouble calls
8. Knowledgebase for information sharing

Clinical Collaboration

Establishing a venue for clinical collaboration is another important, perhaps essential, element of a successful OAN. Even for competing organizations, there is room to collaborate on quality and value alignment care processes. Indeed, the experience in South Carolina has demonstrated that for some services, telehealth can actually help forge partnerships across potentially competing health systems.

This stage was first demonstrated in a partnership between the Medical University of South Carolina (MUSC) and the South Carolina Department of Mental Health. Both entities serve many of the same community hospitals by providing emergency room consultations via telehealth. Although the services and even video end points differed, the two institutions worked together to leverage technology resources and agreed-upon approaches to solving technology problems for the sites they commonly serve. Technical collaboration has matured between the provider groups with common practices both in technology choice and security and network support. Clinical collaboration has been established with one provider group providing care on the inpatient side and another in the emergency department and outpatient setting requiring clinician communication for smooth handoffs.

Administrative Collaboration

An important element for the long-term success of a collaborative OAN effort across health systems is leadership buy-in and administrative inclusion. Establishing a mechanism for discussing the administrative elements of contracting, credentialing, scheduling, and business planning while the clinical and technical work is underway can greatly streamline any difficulties faced and lead to synergies in how each participating health system operates in the telehealth space. If possible, having the telehealth network included within the core strategic initiatives and alignment with the health systems mission is ideal for ensuring long-term success of the OAN.

Stage 2 Example: Telemental Health Services in South Carolina

Stage 2 of the OAN was first demonstrated in a partnership between the MUSC and the South Carolina Department of Mental Health. Both entities serve many of the same community hospitals by providing emergency room consultations via telehealth. Although the services and even video end points differed, the two institutions worked together to leverage technology resources and agreed-upon approaches to solving technology problems for the sites they commonly serve. Technical collaboration has matured between the provider groups, with common practices both in technology choice and security and network support. Clinical collaboration has been established with one provider group providing care on the inpatient side and another in the emergency department and outpatient setting requiring clinician communication for smooth

handoffs. Administrative collaboration has also been ongoing such that leadership finds the relationship so important that this same division of telemental health services that began in community hospitals is now replicated in MUSC's own children's hospital as well.

Stage 3: Functional Clinical Network

The first two stages of OAN maturity deal with technical and collaborative infrastructure. In stage 3, there is a focus on the demonstration of a functional interoperable network, with different institutions providing services through common telehealth end points at a patient site. To make this transition, there is a focus on distributing the established guidelines to focus on training and smooth *launching* or *go live* of services. Finally, quality metrics are tracked to ensure fidelity and encourage ongoing quality improvement. Attention to this ongoing process acknowledges that the technical, clinical, and administrative and achievements of stages 1 and 2 may not be fully matured and require ongoing attention and modifications. Embracing interoperability will likely lead to scaling of functionality to cost while using common end points. It is in stage 3 where these realities are confronted to move toward stage 4 while not sacrificing clinical integrity.

Distribution of Best Practices

For providers at multiple institutions to put the OAN technology to use, there must be a shared understanding of the best practices established by the workgroups during stages 1 and 2. Although all intricacies of the technology choices are not necessarily of utmost importance to this group, a basic understanding of telehealth and how to use and troubleshoot the technology is vital. In addition, when a consensus has been reached to the questions listed in Table 2, a forum for distributing these guidelines to a larger group helps to overcome knowledge barriers. The SCTA has worked through the distribution of instructional information in a variety of ways. Both content advisory and education workgroups exist to help increase general awareness and make more detailed training documents available. These range from a large library of promotional videos to a repertoire of online training modules developed in collaboration with the state's public television and broadcasting agencies and the South Carolina Area Health Education Consortium. Although these materials are in continued development, their use in conjunction with state and regional group training sets the stage for more in-depth on-site training. These program-specific training for providers and support staff provide employees from multiple institutions with a shared understanding before any go live.

Program Go Live

Collaborative use of open access technology by multiple institutions requires substantial coordination and support throughout the process of program development and transition to patient care. This is often best facilitated through dedicated telehealth coordinators or designated contact personnel at each institution. These designees are able to provide a source of continuity between IT, clinical, and administrative teams and host forums for discussion among the institutions. As the training is completed, they are able to host a series of *mock* connections before the rollout of services involving actual patient care and then monitor the first several patient encounters. During this transition period, the coordinators are able to help troubleshoot or find solutions to issues that might otherwise provide a less-than-optimal experience for either the patient or provider.

Quality Metric Tracking and Improvement

As in any process improvement effort, establishing metrics of success that can be tracked over time is essential. Ultimately, monitoring successful uses of common end points into the OAN by multiple institutions is a useful metric, but there are likely others needed to obtain this goal. In that stage, the practical realities that inhibit full deployment of the OAN are addressed or at least acknowledged as areas of optimization. These differences in functionality do not exclude the possibility of interoperability, but the clinical and cost needs may understandably be prioritized over interoperability in stage 3. In this stage, these practical realities are acknowledged, and a course is toward interoperability through process improvement and adjustments to technologies.

Stage 3 Example: Federally Qualified and Rural Health Centers

As the established infrastructure for providing health care for underserved citizens, many of the federally qualified health centers and rural health centers in South Carolina provide examples of stage 3 OAN maturity. These centers are able to use standards-based equipment to allow for connections to services between their own sites and community-based sites such as schools and drug and alcohol rehab facilities and to receive care from distant sites. They have contributed to several of the workgroups tasked with establishing standards and technology, administrative and clinical infrastructure, and the development of specific clinical programs. Their staff members and providers have taken advantage of common training materials and sessions to learn and help establish best practices. On-site training at these sites and those providing care is facilitated by coordinators, ensuring that all involved are ready for an organized go live. Patient care at these sites may utilize different workflows; however, the use of standards-based equipment and common practices, enable efficient use of technology. Finally, quality metrics are beginning to be established at these clinics to track and facilitate the continuous improvement of patient care.

Stage 4: Standard Process Network

Once a use case of two institutions using a common end point has been demonstrated, the transition to stage 4 maturity is underway. In this stage, optimization of the network focuses on the elements that will truly make the OAN add value to the regional care system. In this stage, clinical workflows can be standardized across institutions, and economies of scale can be achieved through administrative and technical innovations.

Establishment of Common Workflows

With the standardization of clinical workflows, there is a common experience of the clinical providers and staff, allowing for an elevation of skill level and generalization of knowledge and competencies. Administrative innovations may include common scheduling portals, standardization of contract

language, and streamlining of credentialing needs. From a technical perspective, optimization may include the use of software-assisted functionality that complies with interoperability standards. Technical response teams are now highly coordinated with open lines of communication.

Shared Administrative and Clinical Resources

A common Web portal for the OAN could be instrumental in bringing many of the aforementioned elements together. However, to remain open access, the costs of the network should remain reasonable for broad participation and not require the use of scaled functionality for all services. Willing partners should not be excluded from the network within the bounds of the agreed-upon standards and compliance with established workflows and procedures. Uniquely, at this level of collaboration, partnerships on service are enabled, in which otherwise competing institutions may find themselves collaborating with shared resources, such as common physician call pools when providers are in short supply.

Payer Advocacy

At the most mature stage of maturity, programs are able to reach a scale in which utilization and quality metrics are well established. After some time, at-scale health outcomes and cost efficiencies are also able to be measured. Multiple provider groups are then able to leverage these data to advocate for improved reimbursement either through legislation or direct negotiation with individual payers. Although some programs may ultimately rely on multiple funding mechanisms to ensure sustainability, a collaborative approach to demonstrating the value of telehealth programs and ultimately improving the reimbursement landscape is beneficial to all involved.

Stage 4 Example: School-Based Telehealth

The South Carolina school-based telehealth program is an illustrative example of a statewide telehealth program that has moved through each OAN stage of maturity. Early in this program's development, a commitment was made to adhere to open access standards when selecting technology, and broadband capability in rural schools was established with the help of the FCC-funded PSPN. As multiple health systems and provider groups became involved, the school-based telehealth workgroup within the SCTA was established to provide a framework for team building and sharing of ideas related to both clinical and technological standards. A collaborative IT support network was essential to the growth of this program as relationship building between school and provider IT teams was fundamental to both establishing and maintaining connectivity. Eventually, it became evident that a tiered call pool approach in which a school nurse's request for a visit could be sent first to a local provider, then to a provider group at the most local regional health care system, and then to a group at the largest state medical system was the best strategy to keep care local while still providing a quick response. For this to become a reality, a common clinical and administrative support network was established. This model has most recently been implemented using a common software platform that can be accessed by multiple end points that meet the open access standards. Although this platform has not become a fully statewide portal, its use has encouraged common workflows and processes, and with increased collaboration, the program has increased efficiency and utilization. Most recently, data have shown that the program may be associated with more cost-effective health care utilization patterns, and the provider groups are working together to advocate for improved reimbursement policies from a variety of payers.

Exception to the Rule: Intentional Fragmentation of an Open Access Network

Interoperability in telehealth solutions is often a goal rather than a full reality for most health systems. Although this is true for many health care IT solutions, it is compounded with telehealth because of the realities of different functionalities needed for different clinical situations. Health systems and telehealth vendors are often in diligent pursuit of an *enterprise solution*, though modifications and exceptions to the standardizations in place remain common for certain clinical services. Embracing the idea that alternative technologies may be needed to achieve clinical goals can be helpful if it allows for a focus on prioritizing what exceptions and options are tolerable, and to work on the integrations needed. This prioritization can be seen as being intentional with your telehealth systems' fragmentation. Areas of prioritization to consider when faced with conflicting telehealth technology needs include the following:

1. Clinical functionality
2. Delays in service delivery pending prolonged integrations
3. Functionalities that add efficiencies not present in a health system's legacy workflows
4. Cost
5. End user experience, including the patient

Once a set of priorities are established, and some fragmentation of the telehealth technology ecosystem is allowed, it should also be the goal of eventual full standardization. Working toward a system with smooth user experience, maximized functionality, and optimal cost may require long-term goal setting and shared decision making across varying levels of health systems governance.

Examples of Intentional Fragmentation From the South Carolina Open Access Network

Example 1 of Deviation From the Open Access Network: Telestroke

Telestroke, which is a highly successful use of telehealth across the country, is an example of successful clinical collaboration on common technology that is not tied to the South Carolina OAN. The operational needs of this type of program include on-demand call access to providers and rapid integration of radiological imaging, which often leads to the use of niche, proprietary software apps, as it did here in South Carolina. This telestroke network has led the way in terms of clinical collaboration, with multiple health systems contributing to a common pool of providers for the good of the region and sustainability of the service. However, integrating the OAN with this high-volume, high-stakes service remains a challenge, and thus intentional fragmentation of the OAN permitted.

Example 2 of Deviation From the Open Access Network: Tele-Intensive Care Unit

One example of a telehealth program that is unlikely to meet even the most basic OAN standards are the most widely used technologies for tele-intensive care units (tele-ICUs). Tele-ICU, as conventionally defined, offers continuous remote patient monitoring from a multidisciplinary clinical ICU team led by an intensivist in a fixed space operations center. Each ICU patient room being monitored is wired with two-way audiovisual communication technology, and an emergency alert button is installed that can be activated during patient crisis. The most common technology for this clinical application includes proprietary software that enables a centralized patient census along with easy access to the audiovisual communication system. The two-way communication system is controlled by the remote clinicians and is supported by a separate vendor from the tele-ICU system. Some clinical information derived from the EHR is included in the centralized system, although this varies across programs and is often limited to basic electronic data provided from discreet data fields in the EHR (eg, patient vital signs and laboratory testing). A tele-ICU clinician workstation in the operations center includes 8 to 12 separate monitors and has numerous technologic interfaces established between the operations center and the remote ICU, including EHR interfaces, physician order entry interfaces, bedside monitor waveform interfaces, radiology viewing interfaces, and internal communication systems for operations center personnel.

Discussion

Telehealth is a rapidly growing component of the US health care system. Yet, to achieve large-scale adoption, policy makers and health care stakeholders should consider options that cross the traditional boundaries of proprietary health care marketplaces. The OAN is one approach that facilitated the statewide telehealth network in South Carolina.

The approach to OAN development described here provides a roadmap for achieving a functional telehealth network across independent health systems. The South Carolina experience reveals both successes and challenges in achieving this goal. The next steps toward the development of OANs would include advocacy and ongoing engagement with the developers of telehealth technologies regarding their commitment to interoperability. In addition, focus on health system partnerships, and collaborative processes round out the essential elements of the OAN.

Acknowledgments

This publication was supported by the Health Resources and Services Administration (HRSA) of the US Department of Health and Human Services (HHS) as part of the National Telehealth Center of Excellence Award (U66 RH31458-01-00). The contents are those of the author(s) and do not necessarily represent the official views of, nor an endorsement, by HRSA, HHS, or the US government.

Conflicts of Interest

None declared.

References

1. Darkins A. The growth of telehealth services in the Veterans Health Administration between 1994 and 2014: a study in the diffusion of innovation. Telemed J E Health 2014 Sep;20(9):761-768. [doi: 10.1089/tmj.2014.0143] [Medline: 25184945]
2. Yang Y, Iqbal U, Ching JH, Ting JB, Chiu H, Tamashiro H, et al. Trends in the growth of literature of telemedicine: A bibliometric analysis. Comput Methods Programs Biomed 2015 Dec;122(3):471-479. [doi: 10.1016/j.cmpb.2015.09.008] [Medline: 26415760]
3. Wade VA, Hamlyn JS. The relationship between telehealth and information technology ranges from that of uneasy bedfellows to creative partnerships. J Telemed Telecare 2013 Oct;19(7):401-404. [doi: 10.1177/1357633X13506533] [Medline: 24218354]
4. President's Council of Advisors on Science and Technology (PCAST). The White House. 2010 Dec. Report to the President Realizing the Full Potential of Health Information Technology to Improve Healthcare for Americans: The Path Forward URL: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/pcast-health-it-report.pdf [accessed 2020-03-17]
5. National Quality Forum. Washington, DC: National Quality Forum; 2017. A Measurement Framework to Assess Nationwide Progress Related to Interoperable Health Information Exchange to Support the National Quality Strategy URL: https://www.qualityforum.org/Publications/2017/08/Creating_a_Framework_to_Support_Measure_Development_for_Telehealth.aspx [accessed 2020-03-17]
6. South Carolina Telehealth Alliance (SCTA). 2019. URL: https://www.sctelehealth.org/ [accessed 2020-03-17]
7. International Telecommunication Union (ITU). 2018. Definition of 'Open Standards' URL: https://www.itu.int/en/ITU-T/ipr/Pages/open.aspx [accessed 2020-03-17]
8. Palmetto State Providers Network. 2019. URL: https://www.pspnsc.org/ [accessed 2020-03-17]

Abbreviations

EHR: electronic health record

FCC: Federal Communications Commission
HHS: Health and Human Services
HRSA: Health Resources and Services Administration
ICU: intensive care unit
IT: information technology
ITU: International Telecommunications Union
MUSC: Medical University of South Carolina
OAN: open access network
PSPN: Palmetto State Providers Network
SCTA: South Carolina Telehealth Alliance
Tele-ICU: tele-intensive care unit

Edited by G Eysenbach; submitted 08.12.19; peer-reviewed by T Aslanidis, S Sarbadhikari; comments to author 22.01.20; revised version received 28.01.20; accepted 29.01.20; published 22.05.20.

Please cite as:
King K, Ford D, Haschker M, Harvey J, Kruis R, McElligott J
Clinical and Technical Considerations of an Open Access Telehealth Network in South Carolina: Definition and Deployment
J Med Internet Res 2020;22(5):e17348
URL: http://www.jmir.org/2020/5/e17348/
doi:10.2196/17348
PMID:32441657

Original Paper

Artificial Intelligence–Assisted System in Postoperative Follow-up of Orthopedic Patients: Exploratory Quantitative and Qualitative Study

Yanyan Bian[1*], MD; Yongbo Xiang[1*], MD; Bingdu Tong[1], MA; Bin Feng[1], MD; Xisheng Weng[1], MD

Department of Orthopedic Surgery, Peking Union Medical College Hospital, Chinese Academy of Medical Science and Peking Union Medical College, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Xisheng Weng, MD
Department of Orthopedic Surgery
Peking Union Medical College Hospital
Chinese Academy of Medical Science and Peking Union Medical College
No 1 Shuaifuyuan, Dongcheng District
Beijing, 100073
China
Phone: 86 13021159994
Email: doctorwxs@163.com

Abstract

Background: Patient follow-up is an essential part of hospital ward management. With the development of deep learning algorithms, individual follow-up assignments might be completed by artificial intelligence (AI). We developed an AI-assisted follow-up conversational agent that can simulate the human voice and select an appropriate follow-up time for quantitative, automatic, and personalized patient follow-up. Patient feedback and voice information could be collected and converted into text data automatically.

Objective: The primary objective of this study was to compare the cost-effectiveness of AI-assisted follow-up to manual follow-up of patients after surgery. The secondary objective was to compare the feedback from AI-assisted follow-up to feedback from manual follow-up.

Methods: The AI-assisted follow-up system was adopted in the Orthopedic Department of Peking Union Medical College Hospital in April 2019. A total of 270 patients were followed up through this system. Prior to that, 2656 patients were followed up by phone calls manually. Patient characteristics, telephone connection rate, follow-up rate, feedback collection rate, time spent, and feedback composition were compared between the two groups of patients.

Results: There was no statistically significant difference in age, gender, or disease between the two groups. There was no significant difference in telephone connection rate (manual: 2478/2656, 93.3%; AI-assisted: 249/270, 92.2%; P=.50) or successful follow-up rate (manual: 2301/2478, 92.9%; AI-assisted: 231/249, 92.8%; P=.96) between the two groups. The time spent on 100 patients in the manual follow-up group was about 9.3 hours. In contrast, the time spent on the AI-assisted follow-up was close to 0 hours. The feedback rate in the AI-assisted follow-up group was higher than that in the manual follow-up group (manual: 68/2656, 2.5%; AI-assisted: 28/270, 10.3%; P<.001). The composition of feedback was different in the two groups. Feedback from the AI-assisted follow-up group mainly included nursing, health education, and hospital environment content, while feedback from the manual follow-up group mostly included medical consultation content.

Conclusions: The effectiveness of AI-assisted follow-up was not inferior to that of manual follow-up. Human resource costs are saved by AI. AI can help obtain comprehensive feedback from patients, although its depth and pertinence of communication need to be improved.

(J Med Internet Res 2020;22(5):e16896) doi:10.2196/16896

KEYWORDS

artificial intelligence; conversational agent; follow-up; cost-effectiveness

Introduction

Artificial intelligence (AI) is a system that can correctly interpret external data, learn from such data, and flexibly apply the acquired knowledge to achieve specific tasks and goals. With the dramatic improvement of computer power to process big data, AI is already ubiquitous in our daily life in the past 50 years. In the past several years, the application prospect of AI in surgery, radiology, dermatology, and oncology is up-and-coming [1-3].

Telemedicine is a medical method used to provide clinical and educational services for remote areas. Information and communication technology are used to transmit medical information [4,5]. Telemedicine attempts to overcome challenges in health services delivery due to distance, time, and rough terrain by improving cost-effectiveness and accessibility of health services in both developing and developed regions [6]. It is an "open and evolving science that incorporates advancements in new technologies and adapts to the changing health demands and social environments." [7] With the growth of AI and big data analytics, the scope and capability of telemedicine have expanded in recent years. Current telemedicine applications can be divided into four categories: patient monitoring, medical information technology, AI-assisted diagnosis, and information analysis collaboration [7]. With the assistance of AI, telemedicine might be an effective method for disease assessment, diagnosis, management, and monitoring [6], especially in chronic disease [8-12], skin diseases [13], and postoperative follow-up care [14].

Postoperative follow-up is an essential part of orthopedic surgery. Medical institutions can provide service for discharged postoperative patients through follow-up. Traditional methods include phone calls, emails, visiting, and reexamination at clinic; all of these methods need a lot of medical resources. With the development of AI and telemedicine, computers or mobile phones could be used to complete more and more of the follow-up work.

In this paper, we describe an AI-assisted system for the follow-up of patients after surgery. The study's aim was to compare the cost-effectiveness and feedback composition of the AI-assisted system with a traditional method for postoperative follow-up.

Methods

Ethics Approval and Consent to Participate

This study was approved by the Research and Ethics Institutional Committee of Peking Union Medical College Hospital (PUMCH).

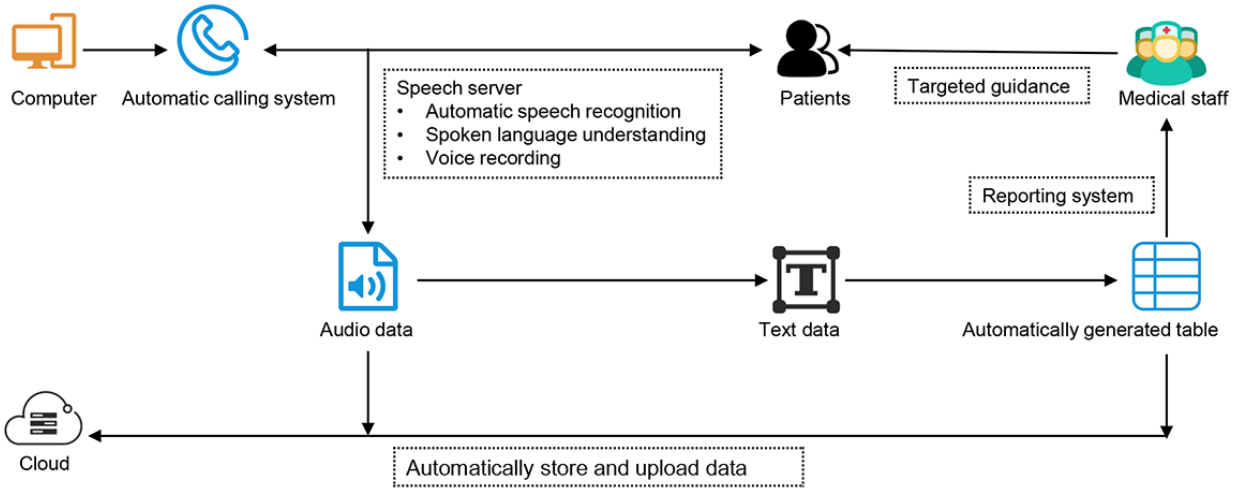
Patient Enrollment and Data Collection

The AI-assisted follow-up system was launched in the Orthopedic Department of PUMCH in 2019. We enrolled 270 patients who had undergone orthopedic surgery at PUMCH from April 2019 to May 2019. This group of patients was defined as the AI-assisted follow-up group since all of the postoperative follow-up was completed by an AI-assisted follow-up system. 2,656 patients who had undergone orthopedic surgery at PUMCH from April 2018 to March 2019 were enrolled as the manual follow-up group. The postoperative follow-up of this group was completed by manual phone calls. Patient characteristics for both groups were collected from the hospital information management system including gender, age, and disease type.

AI-Assisted Follow-Up System

Figure 1 presents a block diagram of the AI-assisted follow-up system. The AI-assisted follow-up system obtained baseline information for each discharged patient including the ID number, gender, age, discharge date, diagnosis, telephone number, and caller location. The system called patients via automated speech telephony delivery in batches from 8:30 AM to 8:30 PM every day allowing hundreds of calls to be made daily. Interactions between the system and patients were based on machine learning, speech recognition, spoken language understanding, and human voice simulation technology. Communication contents included patient satisfaction in the hospital environment, nursing, and health education; wound recovery; functional training; postoperative complications; and other surgery-related medical consulting. The system was able to identify dialects in different parts of China via speech recognition technology and voice information was converted into text in real time. An example of transcribed dialogue that was automatically converted into text is shown in Textbox 1. A report, shown in Table 1, was generated and automatically uploaded to the cloud afterward. Surgeons and nurses could review the report and respond to patient feedback, if necessary.

Figure 1. Block diagram of the AI-assisted follow-up system.



Textbox 1. Transcription of a sample dialogue.

System: Hello, this is Peking union medical college hospital. Could you please take a few minutes to complete a follow-up survey? Thank you!

Patient: ok.

System: our follow-up questions mainly include three aspects: ward environment, propaganda, and nursing. All of the following questions are ranked from 0 to 5, with 5 being the most satisfactory and 1 being the worst. We will record your questions at the end of the follow-up, and would you please help us complete the complete survey?

Patient: yes, no problem.

System: what is your score for the quiet and tidy environment of the ward? The full score is 5 points.

Patient: well, give it a five.

System: ok, may I ask how satisfied you are with the ward environment introduction, health guidance, and the matters needing attention after discharge, etc., given by the nurses during your stay in the hospital? Would you please rate

Patient: well, give it a five.

System: ok, please rate the nurses' responsibility, fit degree, injection technique, regular patrol, response speed, and other nursing work.

Patient: well, five points.

System: ok, what are your complaints about this hospitalization?

Patient: perhaps the most unsatisfactory part is that the doctor is too busy and has too few opportunities to communicate with patients.

System: do you have any other questions for feedback or consultation?

Patient: no.

System: thank you very much for your cooperation. We have recorded your problems and will reply to you as soon as possible. Please keep your mobile phone free and remember to review it on time.

Table 1. Automatically generated table for the dialogue in Textbox 1.

Parameter	Information
Basic information	Patient name, ID number, gender, age, discharge date, diagnosis, and caller location[a]
Dialogue result	System hang up after dialogue
Number of dials	1
Call duration (s)	97
Question 1 score	5
Question 2 score	5
Question 3 score	5
Total score	15
Feedback question	Doctors are too busy to communicate with patients
Follow-up date	2019/6/24

[a]Basic information is automatically extracted from the hospital information management system.

Manual Follow-Up

Follow-up with the control group was performed by phone. Calls were made one by one depending on the availability of human resources. The communication contents were consistent with the AI-assisted follow-up group. Reports were recorded and uploaded manually. Feedback from patients were recorded by the operator and reported to surgeons and nurses.

Evaluation Indicators

Evaluation indicators of follow-up included telephone connection rate, follow-up rate, feedback collection rate, and session duration. Telephone connection rate = (number of effective follow-ups + number of invalid follow-ups) / number of patients called × 100%. Follow-up rate = number of effective follow-ups / (number of effective follow-ups + number of invalid follow-ups) × 100%. Feedback collection rate = number of patients with effective feedback / number of effective follow-ups × 100%. Effective follow-up was defined as the complete collection of data in Table 1 (excluding the feedback question parameter), while invalid follow-up was defined as the absence or incomplete collection of data in Table 1. Effective feedback included patient feedback about nursing, health education, hospital environment, and medical consulting. Moreover, feedback content should be specific and constructive such as dissatisfaction with hospital food or postoperative wound rupture.

Statistical Methods

Data analysis was performed using SPSS statistical software (version 23.0, IBM Inc). Pearson chi-square test or Fisher exact test was used to compare categorical variables (age, gender, medical condition, telephone connection rate, follow-up rate, and feedback rate). The normality of continuous variables (session duration) was tested with a modified Kolmogorov-Smirnov test. Unpaired *t* tests were performed for those following a normal distribution. All tests were two-sided. Data were considered to be statistically significant for $P<.05$.

Results

Patient Characteristics

Patient characteristics for the two groups are shown in Table 2. Group differences in age, gender, and disease category were not statistically significant (Table 2).

Table 2. Comparison of patient characteristics in manual and artificial intelligence–assisted follow-up groups.

Characteristics	Manual follow-up group, n (%)	AI-assisted[a] follow-up group, n (%)	Chi-square (*df*)	*P* value
Number of patients	2656 (100)	270 (100)		
Age (years)			0.8 (3)	.86
<50	805 (30.3)	77 (28.5)		
50-59	480 (18.1)	47 (17.4)		
60-69	861 (32.4)	94 (34.8)		
≥70	510 (19.2)	52 (19.3)		
Gender			1.7 (1)	.19
Male	994 (37.4)	112 (41.5)		
Female	1662 (62.6)	158 (58.5)		
Disease category			4.1 (2)	.13
Spinal disease	1676 (63.1)	154 (57.0)		
Joint disease	938 (35.3)	110 (40.7)		
Other	42 (1.6)	6 (2.2)		

[a]AI-assisted: artificial intelligence–assisted.

Cost-Effectiveness

As shown in Table 3, there was no significant difference in either telephone connection rate (manual: 2478/2656, 93.3%; AI-assisted: 249/270, 92.2%; *P*=.50) or follow-up rate (manual: 2301/2478, 92.9%; AI-assisted: 231/249, 92.8%; *P*=.96) between two groups. However, the feedback collection rate in the AI-assisted follow-up group was significantly higher than that in the manual follow-up group (manual: 68/2656, 2.5%; AI-assisted: 28/270, 10.3%; *P*<.001). The approximate session duration of the manual follow-up group varied from 60 seconds to 180 seconds, according to interviews with the operators. An extra 120 to 180 seconds were also necessary for material recording and uploading. Therefore, the average time spent on each patient in the manual follow-up group was approximately 3-6 minutes. We recorded the total time that the operators spent on randomly following 100 patients (9.3 hours). The average session duration of the AI-assisted follow-up group was 87.7 (SD 39.5) seconds. However, none of this time required human resources, and the AI system generated data reports automatically. In this way, the time spent on AI-assisted follow-up was close to 0 hours. Compared with manual follow-up, the AI-assisted follow-up were of shorter duration.

Table 3. Comparison of manual and artificial intelligence–assisted follow-up indicators.

Indicators	Manual follow-up	AI-assisted[a] follow-up	Chi-square (*df*)	*P* value
Telephone connection			0.4 (1)	.50
Number of effective follow-ups + invalid follow-ups	2478	249		
Number of patients called	2656	270		
Rate, %	93.3	92.2		
Follow-up			0.003 (1)	.96
Number of effective follow-ups + invalid follow-ups	2301	231		
Number of effective follow-ups + invalid follow-ups	2478	249		
Rate, %	92.9	92.8		
Feedback collection			47.1 (1)	<.001
Number of patients with effective feedback	68	28		
Number of effective follow-ups + invalid follow-ups	2656	270		
Rate, %	2.5	10.3		
Time spent, hours per 100 patients	9.3	0	N/A[b]	N/A

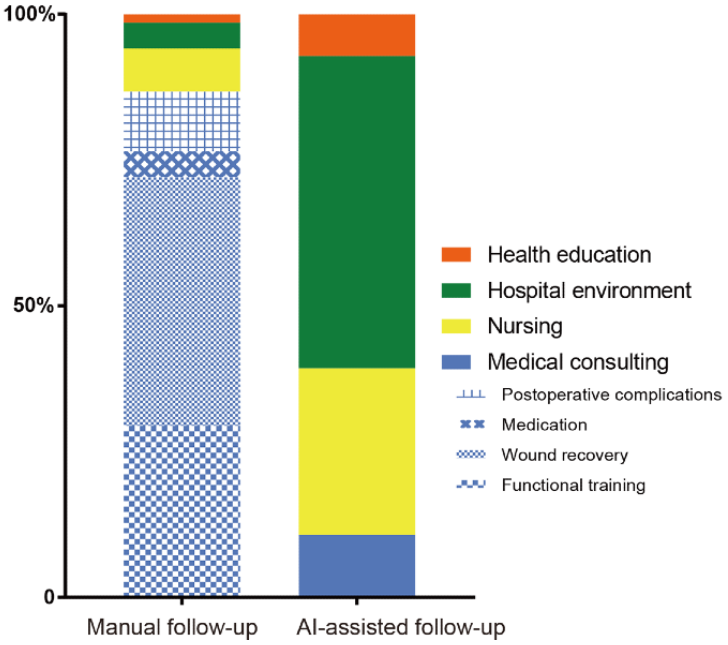
[a]AI-assisted: artificial intelligence–assisted.

[b]N/A: not applicable.

Feedback Composition

The composition of feedback content is shown in Figure 2. In the manual follow-up group, 87.0% of the feedback were medical consultation–related, including functional training, wound recovery, medication usage and postoperative complication (29.5%, 42.7%, 4.5%, and 10.3%, respectively). Feedback related to nursing, health education and hospital environment accounted 7.3%, 1.4%, and 4.3%, respectively. In the AI-assisted follow-up group, only 10.7% of the feedback were related to medical consultation and most of the feedback focused on hospital environment, nursing, and health education (53.6%, 28.6%, and 7.1%, respectively).

Figure 2. Composition of feedbacks in AI-assisted follow-up group and manual follow-up group.



Discussion

Principal Findings

In this paper, we put forward a new methodology for postoperative data collection for follow-up of patients who had undergone orthopedic surgery. The AI-assisted follow-up system is intended to facilitate follow-up efficiency via machine learning, speech recognition, and human voice simulation technology.

In the cost-effectiveness analysis, our study found that there was no significant difference in the telephone connection rate and follow-up rate between the two groups, suggesting the effectiveness of AI was not inferior to the traditional manual method. AI-assisted follow-up could replace traditional manual follow-up to some extent. At PUMCH, follow-up was traditionally completed by nurses or telephone operators where the average time spent on each patient was about 3-6 minutes. The AI-assisted system, however, was able to follow up 5-7 patients simultaneously using telephone relay technology. Thousands of patients could be followed up daily while the time spent on AI-assisted follow-up was close to 0 hours, saving a lot of manpower and human resources.

AI-assisted follow-up did improve the feedback rate, but the composition of feedback was different between the two groups. Feedback from the AI-assisted follow-up group were mainly related to nursing, health education, and hospital environment; only 11% was related to medical consulting. Conversely, this number was 87% in the manual follow-up group. After interviewing the operators, we found a possible explanation to be the telephone operator (usually nurses) were more likely to record feedback that was difficult for them to respond directly, such as medical consultation. In order to improve follow-up efficiency, they tended to respond directly to feedback related to nursing, hospital environment, and health education. This feedback would not have been recorded in the follow-up materials. It also explains why the feedback rate was lower in the manual follow-up group. Another possible explanation was that compared with AI-assisted follow-up, patients could communicate more naturally and deeply with operators. As a result, they may have been more willing to put forward professional medical consultation feedback to operators. In the manual follow-up group, medical consulting could be divided into four categories while in the AI-assisted follow-up group, medical consultation was difficult classify because it was not pertinent.

Limitations

Our work has several limitations. First, apart from phone calls, there are other communication methods that could be combined with AI, such as text messages, computer software, and smartphone apps. Anthony et al [15] invented an automated mobile phone messaging platform for orthopedic trauma patients that improved the responding rate after trauma procedures; however, many elderly patients were not familiar with texting. Therefore, texting might not be suitable for follow-up. Another option is the chatbot. It is a computer program or smartphone app based on AI that can communicate with people via auditory or text [15]. Medical chatbots have been used in disease diagnosis [16], management [17], and monitoring [18]. Recent research showed that chatbots were a convenient method to help

patients address common concerns after ureteroscopy [19]. Integrating chatbots into the telemedicine system could be used to help assess disease conditions and provide self-care recommendations for patients [20]. Therefore, future work may include developing a chatbot software or app for additional medical purposes and to provide better personal service.

The second limitation of this study is that the probation period of the AI-assisted follow-up system is not too long. It is generally believed that with the increasing of machine learning time, the AI-assisted system would become more intelligent. Therefore, we should pay more attention to this system in the future.

Conclusions

In this research, we found that the effectiveness of AI-assisted follow-up was not inferior to the manual follow-up. Moreover, human resources costs could be saved with the assistance of artificial intelligence. Compared with manual follow-up, AI-assisted follow-up obtains more comprehensive feedback, but feedback lacks depth and pertinence. Therefore, the application of an AI-assisted follow-up system in hospital ward management has the potential to improve telemedicine follow-up service and patient satisfaction.

Acknowledgments

The AI-assisted follow-up system was designed by the Beijing Qiyun Zhida Technology Co, Ltd. This system is currently used in the field of trade, finance, and medicine.

Conflicts of Interest

None declared.

References

1. Bouletreau P, Makaremi M, Ibrahim B, Louvrier A, Sigaux N. Artificial Intelligence: Applications in orthognathic surgery. J Stomatol Oral Maxillofac Surg 2019 Sep;120(4):347-354. [doi: 10.1016/j.jormas.2019.06.001] [Medline: 31254637]
2. Smith KC. A discrete speech recognition system for dermatology: 8 years of daily experience in a medical dermatology office. Semin Cutan Med Surg 2002 Sep;21(3):205-208. [doi: 10.1053/sder.2002.34950] [Medline: 12322994]
3. Garg S, Williams NL, Ip A, Dicker AP. Clinical Integration of Digital Solutions in Health Care: An Overview of the Current Landscape of Digital Technologies in Cancer Care. JCO Clin Cancer Inform 2018 Dec;2:1-9 [FREE Full text] [doi: 10.1200/CCI.17.00159] [Medline: 30652580]
4. WHO. A Health Telematics Policy in Support of WHO'S Health-For-All Strategy for Global Development. In: World Health Organization. 1997 Presented at: Report of the WHO Group Consultation on Health Telematics; 11-16 December; Geneva.
5. Dinya E, Tamás T. Health informatics: eHealth and Telemedicine. 2013 Presented at: Semmelweis University Institute of Health Informatics; 2013; Budapest.
6. Kuziemsky C, Maeder AJ, John O, Gogia SB, Basu A, Meher S, et al. Role of Artificial Intelligence within the Telehealth Domain. Yearb Med Inform 2019 Aug;28(1):35-40 [FREE Full text] [doi: 10.1055/s-0039-1677897] [Medline: 31022750]
7. Ryu S. Telemedicine: Opportunities and Developments in Member States: Report on the Second Global Survey on eHealth 2009 (Global Observatory for eHealth Series, Volume 2). Healthc Inform Res 2012;18(2):153. [doi: 10.4258/hir.2012.18.2.153]
8. Taylor Y, Eliasson A, Andrada T, Kristo D, Howard R. The role of telemedicine in CPAP compliance for patients with obstructive sleep apnea syndrome. Sleep Breath 2006 Sep;10(3):132-138. [doi: 10.1007/s11325-006-0059-9] [Medline: 16565867]
9. Montani S, Bellazzi R, Larizza C, Riva A, d'Annunzio G, Fiocchi S, et al. Protocol-based reasoning in diabetic patient management. Int J Med Inform 1999 Jan;53(1):61-77. [doi: 10.1016/s1386-5056(98)00103-8] [Medline: 10075131]
10. Whitlock W, Brown A, Moore K, Pavliscsak H, Dingbaum A, Lacefield D, et al. Telemedicine improved diabetic management. Military medicine 2000;165(8):579-584. [doi: 10.1093/milmed/165.8.579]
11. Morlion B, Verbandt Y, Paiva M, Estenne M, Michils A, Sandron P, et al. A telemanagement system for home follow-up of respiratory patients. IEEE Eng Med Biol Mag 1999;18(4):71-79. [doi: 10.1109/51.775491] [Medline: 10429904]
12. Kinne G, Droste C, Fahrenberg J, Roskamm H. Symptomatic myocardial ischemia and everyday life: implications for clinical use of interactive monitoring. J Psychosom Res 1999 Apr;46(4):369-377. [doi: 10.1016/s0022-3999(98)00096-8] [Medline: 10340236]
13. Elsner P, Bauer A, Diepgen TL, Drexler H, Fartasch M, John SM, et al. Position paper: Telemedicine in occupational dermatology - current status and perspectives. J Dtsch Dermatol Ges 2018 Aug;16(8):969-974. [doi: 10.1111/ddg.13605] [Medline: 29998596]
14. Young K, Gupta A, Palacios R. Impact of Telemedicine in Pediatric Postoperative Care. Telemed J E Health 2019 Nov;25(11):1083-1089. [doi: 10.1089/tmj.2018.0246] [Medline: 30517056]

15. Anthony CA, Volkmar A, Shah AS, Willey M, Karam M, Marsh JL. Communication with Orthopedic Trauma Patients via an Automated Mobile Phone Messaging Robot. Telemed J E Health 2018 Jul;24(7):504-509. [doi: 10.1089/tmj.2017.0188] [Medline: 29261036]
16. Ni L, Lu C, Liu N, Liu J. Towards a smart primary care chatbot application. 2017 Presented at: In International Symposium on Knowledge and Systems Sciences (pp.). Springer; 2017; Singapore p. 38-52. [doi: 10.1007/978-981-10-6989-5_4]
17. Chung K, Park RC. Chatbot-based heathcare service with a knowledge base for cloud computing. Cluster Comput 2018 Mar 16;22(S1):1925-1937. [doi: 10.1007/s10586-018-2334-5]
18. Piau A, Crissey R, Brechemier D, Balardy L, Nourhashemi F. A smartphone Chatbot application to optimize monitoring of older patients with cancer. Int J Med Inform 2019 Aug;128:18-23. [doi: 10.1016/j.ijmedinf.2019.05.013] [Medline: 31160007]
19. Goldenthal SB, Portney D, Steppe E, Ghani K, Ellimoottil C. Assessing the feasibility of a chatbot after ureteroscopy. Mhealth 2019;5:8 [FREE Full text] [doi: 10.21037/mhealth.2019.03.01] [Medline: 31019961]
20. Fadhil A. Beyond patient monitoring: Conversational agents role in telemedicine & healthcare support for home-living elderly individuals. arXiv preprint arXiv.0 2018:6000.

Edited by G Eysenbach; submitted 07.11.19; peer-reviewed by V Khanna, D Pförringer; comments to author 10.02.20; revised version received 06.04.20; accepted 08.04.20; published 26.05.20.

Please cite as:
Bian Y, Xiang Y, Tong B, Feng B, Weng X
Artificial Intelligence–Assisted System in Postoperative Follow-up of Orthopedic Patients: Exploratory Quantitative and Qualitative Study
J Med Internet Res 2020;22(5):e16896
URL: http://www.jmir.org/2020/5/e16896/
doi:10.2196/16896
PMID:32452807

Original Paper

Internet Narratives Focused on Health Travelers' Experiences in India: Qualitative Analysis

Joseph Brown[1], BA, MSA, DHA; James Johnson[1*], PhD, MPA, MSc; Margaret E Ozan-Rafferty[2*], RN, MHA, MBA, DHA; Manoj Sharma[3*], MD; Salvatore Barbera[4*], BA, MS, DHA

[1]The Herbert H & Grace A Dow College of Health Professions, Central Michigan University, Mount Pleasant, MI, United States

[2]Blessing Health System, Quincy, IL, United States

[3]Covenant HealthCare, Saginaw, MI, United States

[4]Health Administration Management Program, College of Business, Florida Atlantic University, Boca Raton, FL, United States

[*]these authors contributed equally

Corresponding Author:
Joseph Brown, BA, MSA, DHA
The Herbert H & Grace A Dow College of Health Professions
Central Michigan University
1280 E Campus Dr
Mount Pleasant, MI, 48859
United States
Phone: 1 9897744000
Email: brown7jc@cmich.edu

Abstract

Background: The medical tourism industry is currently popular in India, but there is no confirmation of the common perspectives among the country's medical travelers.

Objective: This qualitative research study analyzed web-based narratives from health travelers visiting India and described the themes of their experiences. This study aimed to answer the following primary question: What can we learn about health travelers' experiences in India from an analysis of their web-based narratives? The secondary questions were as follows: (1) What are the primary health care reasons for which patients in the examined narratives traveled to India? (2) What can be derived from the narratives regarding medical tourists' satisfaction with the outcome and result of the treatment they received in India? (3) What are some positive and negative factors influencing medical tourists' perceptions and overall experiences about their health travel to India? (4) What are the characteristics of medical tourists who write web-based narratives regarding their health experiences in India?

Methods: Publicly available narratives written by medical tourists who visited India were obtained from a Google search. The narratives included blog posts and discussion board posts by medical tourists. The analysis process consisted of initial open coding being conducted on the narratives to create initial codes and identify common themes with a focus on the primary research question and subquestions.

Results: Although Mumbai, Chennai, and New Delhi were not the only destination cities mentioned, these were the most popular cities patients visited for care. The medical tourists, who stated their origin country, came from one of the following continents: Africa, Europe, North America, and Oceania. Dental care, Ayurveda treatment, and eye care were the most popular types of care that medical tourists sought. The results showed that most of the medical tourists were happy with the overall experience of receiving care in India. The most popular themes with regard to the patients' satisfaction were low costs, good customer service, and services being offered that were unavailable in their home country. When negative feedback was provided, it was mainly concentrated on the overall environment of India being unorganized and unsanitary.

Conclusions: Primarily, the study's findings can benefit health care providers and patients. Providers hosting medical tourists in India can use negative feedback to improve their services; similarly, providers who are losing patients to medical tourism can identify opportunities for improvement (ie, why are we losing patients). Indian providers hosting medical tourists should keep their prices competitive and continue to provide exceptional service; however, they should do their best to lessen the crowdedness of their facilities while making sure they are esthetically pleasing. Providers losing patients to medical tourism need to identify

ways to ensure their services match the benefits that their international counterparts are providing, such as competitive pricing and expansion on the services provided.

(J Med Internet Res 2020;22(5):e15665) doi:10.2196/15665

KEYWORDS

medical tourism; India; global health; personal narratives; travel-related illness; qualitative research; patient satisfaction; delivery of health care; travel; data collection

Introduction

Background

Depending on the source, the terms "health tourism," "medical tourism," and "wellness tourism" are used very loosely and unsystematically [1]. According to Connell [2], this is most likely because the boundaries between these terms are unclear. For the purposes of this paper, the terms medical tourism and health traveler are used to define instances when a patient travels across international borders to obtain medical care. These medical tourism treatments may range from highly invasive heart surgeries to less invasive procedures such as dental work. Mutalib et al [3] stated that "medical tourism (specifically) is not a new phenomenon, what is new is actually the trend of practicing medical tourism." Lunt et al [4] explained that availability, cost, expertise, and cultural and familial reasons motivate patients to seek care abroad. The majority of medical education in India is subsidized by the government; therefore, the doctors do not have the heavy burden of student loans [5]. This helps to drive down the cost of care. Similarly, Alsharif et al [6] examined medical tourists visiting India, China, Jordan, and the United Arab Emirates and found that the most important reasons for patient travel were cost, provider's reputation, and hospital accreditation. Throughout the paper, the terms posts, narratives, and journals have been used interchangeably to represent the qualitative data analyzed to execute this research.

The internet can play a role in a prospective patient's medical tourism decisions. According to Lunt et al [4], patients consult a range of information sources before making decisions, and the internet plays a key role in providing information from informal networks. Consumers participating in web-based health care communities often reveal personal information in intricate detail [7]. As it relates to health care specifically, trust in the destination assures medical tourists who choose to visit a particular destination that the services provided will be transparent, reliable, and risk and hassle free [8]. According to Lunt et al [4], medical tourists often pay more attention to soft information than hard clinical information.

Prior Work

This study originated from the recommendations in the further research section in the study by Ozan-Rafferty et al [9] titled "A qualitative analysis of internet narratives by health travelers to Turkey". Her study analyzed web-based narratives from health travelers to Turkey and described the themes of their experiences. This study will mimic the established method of the abovementioned study but focus on India instead of Turkey.

Setting

India is roughly one-third the size of the United States but its population is about four times that of the United States [10]. India's private sector providers, along with independent medical tourist–facilitating organizations, have taken the lead in developing medical tourism [11]. The combination of India's economic liberalization, growing middle class, and the rise in medical tourism has contributed to the growth in the number of private, for-profit hospitals [12]. The primary reason for India's growing medical tourism sector is cost savings. On average, medical services provided in India cost almost half the price of those offered by any other developed country [13]. The Indian government has encouraged the growth of medical tourism by issuing special visas for medical tourists [6]. In addition, many patients are intrigued by the possibility of combining their medical treatment with a vacation [2].

There are no authoritative data on the number and flow of medical tourists between nations and continents [14]. This is mainly because there are no international standards for the transmission of information to clinicians in the patient's country of origin [15]. Bernasek [16] cited the World Health Organization's projection that 16 million Americans would travel abroad for medical care in 2017. In their study on medical tourism growing in the Indian health care market, Gupta et al [17] noted that India's Ministry of Tourism has even conducted road shows in West Asia (Dubai, Riyadh, Kuwait, and Doha) to promote medical tourism.

Similar to the research by Ozan-Rafferty et al [9], this research assumes that understanding the experiences of medical tourists will benefit the providers who service current and future medical tourists. In addition, the findings from this study can be used to assist in decision making for patients considering health travel in the future, and they may strengthen health administration education by providing insights on medical tourism [9]. In addition, this study's outcomes may lead to the development of more online communities focused on improving medical tourism. In the future, this study can help further the push of establishing a global tracking system for statistics on the number of medical tourists. More importantly, we hope that this study will encourage more research on the topic of medical tourism by way of its recommendations and drive competition resulting in a better health experience for all patients irrespective of their location.

No funding was required for this study. As per Central Michigan University's Institutional Review Board (2017), it has been determined that this project does not meet the definition of human subject research under the purview of the institutional review board (ie, exempt status). This is because no human

subjects were involved in the research, and no identifiable information was included.

Research Questions

The medical tourism industry is currently popular in India, but there is no confirmation of the common perspectives among the country's medical travelers. The questions to be answered are as follows: (1) What are the primary health care reasons for which patients in the examined narratives traveled to India? (2) What can be derived from the narratives regarding medical tourists' satisfaction with the outcome and the result of the treatment that they received in India? (3) What are some positive and negative factors influencing medical tourists' perceptions and overall experiences about their health travel to India? (4) What are the characteristics of medical tourists who write web-based narratives regarding their health experiences in India?

Methods

Recruitment

This is a qualitative research study based on a narrative analysis from Creswell's qualitative approach [18]. Narratives written by medical tourists visiting India were obtained using publicly available webpages identified via a Google search. The narratives included blog posts and discussion board posts by medical tourists who visited India for care. Data were extracted using purposeful sampling, which focuses on information-rich cases related to a specific topic [19]. Similarly, Blumberg et al [20] stated that purposeful sampling is when researchers select participants arbitrarily for their unique characteristics or their experiences. In this case, the focus was on patients who left their home countries and received care in India. The search terms used included, but were not limited to, "health travel," "medical tourism," "plastic surgery," and "wellness travel" and incorporated "India," "Mumbai," "New Delhi," and "Kolkata." Specific clinical procedures used in the search were determined based on the reviewing sites that promoted health travel to India, as well as findings from the literature review, and terms noted in initial findings.

Results from the first 10 pages of the general Google search were reviewed to find narratives that met this study's criteria, which were as follows: (1) a first-person narrative written by an individual who underwent treatment or their partner on the trip, (2) a narrative written in English, (3) a narrative that includes a description of the type of procedure, (4) a narrative that includes a personal experience of health travel to India, and (5) a publicly available narrative that does not require a password or discussion board membership to be accessed [9]. Then, following the core of the process established in the "A qualitative analysis of internet narratives by health travelers to Turkey" study, the next step limited the search results to only include discussions. Next, results from the first 10 pages of the "discussion" were only filtered results that were reviewed to determine whether there were any eligible narratives meeting the same 5 requirements listed earlier.

On the basis of the research by Ozan-Rafferty et al [9], the following review attributes were considered when collecting the narratives. Any narrative that was longer than a single sentence was added to the queue for analysis. However, some of the narratives were in a chronological story format, and some were shorter in length and outlined a specific part of a general larger experience [21]. Even though posts and blogs related to the topic of the author's health travel were included in the research, other topics written by the author were reviewed to obtain demographic or characteristic information such as the author's age, gender, or country of origin. Multimedia Appendix 1 displays the steps in the narrative process by way of a flow diagram.

On the basis of the types of medical tourism treatments identified in the research review, 256 different keyword searches were performed. These produced more than 800 million search results. Of course, this is when counting results falls outside of the first 10 pages. After the filtering process was completed, there were 53 narratives available for analysis.

Coding Analysis

For the coding process, the identified narratives were copied from the internet and posted in individual Microsoft Word (Microsoft Corporation) documents. Each document was named by using an identifier (ie, "000") that coincides with its entry number in the process and findings log. Similar to the study by Ozan-Rafferty et al [9], the narratives that included more than one entry were collected chronologically from the oldest to the newest. All the narratives combined totaled to 121 pages. The largest page count of a single narrative used was 23. There were 34 narratives that were 1 page each. As each narrative was contained in its own document, the page count was rounded off. For example, a narrative that was one and half pages long was counted as being 2 pages long.

Initial open coding was conducted on the narratives to create initial codes and identify common themes with a focus on the primary research question and subquestions [9]. The unit of reference was a sentence [21,22], and a total of 2536 units of measure were analyzed. In addition, the content of each sentence was examined to identify positive and negative opinions [22]. Written words were analyzed using narrative analysis theory principles to create primary (parent) and secondary (child) themes [9]. As the analysis occurred, notes were taken to clarify the concept behind some of the codes and sentences being analyzed. More than 40 primary themes were identified, and if necessary, memos were kept for each sentence coded to add context and keep track of concepts.

A second review was completed, and the narratives were entered into NVivo Pro Version 11 (QSR International) where the initial codes, key concepts, and themes identified in the open coding were clustered [23]. As was performed in the study by Ozan-Rafferty et al [9], NVivo was used to identify relationships with the various initial themes. Each sentence along with its corresponding code was entered into NVivo. The codes were clustered into categories, and various clusters were used to identify trends and relationships.

A final round of selective coding [23] was completed by scanning all the data and codes and reviewing all themes. Multimedia Appendix 2 displays the theme frequency count table. To confirm coding accuracy, an additional researcher was

approved by the dissertation committee, and he randomly selected 8 of the 53 narratives to review. For the random selection process, the narratives were numbered 1 to 53 but were in no specific order. The 8 narratives selected were independently coded by the additional researcher. After comparing the additional researcher's codes with the primary researcher's initial codes, there was only 96.5% reliability (reliability=number of agreements divided by the number of agreements + disagreements). It is important to note that any personal identification details of the narrator were removed and not considered in this research.

As has been mentioned, there was a committee outside of the additional researcher. This committee consisted of 3 members, and they approved all major components of the study. Given that she had prepared a similar project focusing on Turkey, Dr Ozan-Rafferty was the committee member who assured that the methods were followed correctly. Dr Sharma, originally from India, assured the accuracy of information depicted in the narratives, and Dr Johnson was the chairperson overseeing all aspects of the project.

Data Exclusion

During the search process, it was not very difficult to find weblinks in which patients were discussing their medical experiences in India. However, most of the initial findings were not eligible for this study because they were being posted for promotional purposes. For example, the direct quotes were on the webpages of hospitals that wanted people to come to their facilities for a certain procedure. The articles that quoted medical tourists but were not written by the actual medical tourists or by someone who went on the trip with them were excluded. There were many photo blogs and video blogs (vlogs); by strictly following the method established in previous research on this topic, this type of media was not considered. Adding a photo, video, or the aspect of judging a person's tone could have easily distracted from the overall word analyzation, which was the primary focus of this project. It was not uncommon to find blogs or articles written by doctors who treat medical tourists. If joining a forum or discussion group was required to obtain any narratives, obtaining membership in the discussion forum or discussion group was not done to ensure all information used is truly publicly available. A log of each narrative was created and included the web address of each post. Some medical tourists posted the exact copy of the narrative on multiple webpages, and those duplicates were eliminated from this research.

Results

Overarching Narrative Details

The entire search process took place from November 5 to 16, 2017. Narratives were reread on the web on January 8, 2018, to confirm no additional postings were added. Some search inputs garnered millions of results. Of the narratives that were used, most came from discussion forums where someone would start a thread such as, "Has anyone been to India for…," and previous medical tourists responded by sharing their stories and helping out other prospective health travelers by answering general questions. There were several blogs as well. Some of the blogs were very professional-looking (with photos and embedded videos), whereas others were simple text entries. From time to time, a narrative was found in the comments of an article about medical tourism, so it was important to read all the comments on an identified weblink. At the conclusion of the search, a total of 53 narratives were deemed eligible for analysis. Of those 53 narratives, 6 were blog posts. The remaining 47 narratives were from a discussion thread.

Of the 53 narratives collected, there were only 8 narratives in which the medical tourists' countries of origin (home countries) were unidentifiable. Table 1 displays the continent of origin and country of origin for each patient. This table is in an alphabetical descending order. With 11 narratives, the United States was the country with the highest number of medical tourists visiting India, followed by the United Kingdom and Australia (10 each). There were 3 countries with only 1 medical tourist visiting India: Morocco, Austria, and France.

Table 1. Health travelers' country of origin.

Country	Travelers, n
Africa	**1**
Morocco	1
Europe	**16**
Austria	1
France	1
Germany	2
Ireland	2
United Kingdom	10
North America	**18**
Canada	7
United States	11
Oceania	**10**
Australia	10
Undetermined	**8**
Undetermined	8
Total	53

Narrative Content Details

Of the narratives analyzed, the most frequent procedure availed by a medical tourist was some kind of dental care. This was followed by Ayurveda treatment and Lasik eye surgery. There were a number of services that were only mentioned once across all the narratives. These services ranged from gallbladder surgery to a hair transplant. Although it did occur, it was not a common trend for patients to avail more than one medical service while in India. Table 2 consists of the detailed counts of how often a procedure was availed.

Tables 3-6 provide details regarding each patient and demographic details about where the procedures occurred. With regard to gender, 21 of the medical tourists were men and 18 were women; the gender of the remaining 14 tourists was unidentifiable. In most cases (32 narratives), there was no indicator of the patient's age. Regarding the identified ages, it is notable that 14 of the patients were older than 40 years when their initial post was made. Then, 16 of the narratives referred to a person accompanying the medical tourist while in India, and of those 16, 4 were authored by the person who accompanied the patient. Only 4 narratives mentioned that the patients used a medical tourism–facilitating organization.

Most of the narrators (n=30) only made a single post, and 11 made two posts. One narrator made 21 entries, and another made 39 entries. There were 4 narratives in which years passed between the first and last posts. However, all the other narratives consisted of posts that occurred within a year or another. In addition, 45 of the posts were made between 2010 and 2017. The oldest post analyzed was from March 2005, and the newest post was from December 2017. A total of 29 narratives mentioned the actual treatment date.

The destination city was noted in 38 of the narratives. Mumbai, Chennai, and New Delhi were the most popular cities that patients visited for care. Regarding specific facilities, 23 different facilities were listed, but only 2 were listed more than once: Asian Joint Reconstruction Institute of Chennai and MY EYE.

The authors were very open about their experiences. They shared details such as the food they ate all the way up to their personal fears (such as flying). It was similar to reading a personal journal. From a personality standpoint, very few users described themselves, but when they did, they included descriptors such as humble, self-conscious, and focused on health. Most narrators had a positive attitude rather than a negative one. This is because they stated things such as "I'm excited about the trip" or "I had been looking forward to this." Many sentences that were coded described the narrator as not having a complaining attitude, having an overall positive view on life, or being optimistic about their health's improvement after receiving treatment in India. Contributors also encouraged others. For example, a patient with Crohn disease sent this message to another person diagnosed with Crohn disease: "I really hope she feels better, and I know sometimes it can feel like there is no hope and you're willing to try everything to help alleviate the symptoms." The following sections are reasons for medical tourism, patient satisfaction, and perception influencers. These are used to highlight the significant themes identified in the narratives.

Table 2. Health travelers' primary and additional procedures.

Row labels	Count of primary procedure[a]	Count of additional procedures
Ayurveda	9	N/A[b]
Dental	13	N/A
Eyelid surgery	1	2
Gallbladder surgery	1	3
Gynecomastia surgery	1	N/A
Hair transplant	1	N/A
Heart surgery	1	N/A
Hernia surgery	1	N/A
Joint replacement	3	N/A
Lasik eye surgery	8	2
Limb lengthening	3	N/A
Physical exam	2	N/A
Physiotherapy	1	N/A
Plastic surgery	3	1
Sex reassignment surgery	1	N/A
Shot	1	N/A
Spine surgery	1	N/A
Surgery	1	N/A
Surrogacy	1	N/A
Grand total	53	8

[a]Primary procedure is the procedure the narrator noted first or mentioned as their primary reason for traveling to India for care.

[b]N/A: not applicable.

Table 3. Health travelers' date of initial narrative, demographics, location of treatment, date of treatment, facilitator, facility, and accompanying person.

Author	First post	Gender	Age of patient when post was made (years)	City of procedure	Treatment date	Facilitator	Facility	Accompanying person
001	August 15, 2016	Female	N/A[a]	N/A	N/A	N/A	N/A	N/A
002	November 11, 2010	Male	N/A	New Delhi	2010	N/A	N/A	N/A
003	March 29, 2017	Male	60	Chennai	January 16	N/A	Asian Joint Rebuilding Institute in Chennai	N/A
004	N/A	Female	N/A	Goa	N/A	N/A	Vrundavan Hospital	Spouse
005	July 28, 2010	Male	50-64	Delhi	February 12	SCODE	N/A	N/A
006	July 10, 2016	Female	N/A	New Delhi	N/A	N/A	Delhi Dental Center	N/A
007	January 17, 2010	Female	50-64	Margao	2009	Apollo Victor	Apollo Victor	Daughter
008	April 2, 2016	N/A	N/A	Hyderabad	N/A	N/A	Dr. Motiwala Dental Clinic	N/A
009	June 11, 2016	Female	N/A	Hyderabad	N/A	N/A	N/A	N/A
010	May 20, 2015	Female	50-64	Hyderabad	2014	N/A	Dr. Motiwala Dental Clinic & Implant Center	N/A
011	December 4, 2017	N/A	N/A	Chennai	2017	N/A	Rajan Dental	Husband
012	April 15, 2013	Female	50-64	Nuvem	December 11	N/A	MY EYE	N/A
013	November 1, 2013	Female	35-49	Mapusa	N/A	N/A	N/A	N/A
014	November 5, 2013	Male	50-64	Goa	November 4, 2013	N/A	N/A	N/A
015	December 10, 2014	N/A	N/A	N/A	2015	N/A	N/A	N/A
016	January 1, 2008	N/A	N/A	Chennai	February 27, 2008	N/A	Sankara Nethralaya Hospital	N/A

[a]N/A: not applicable.

Table 4. Health travelers' date of initial narrative, demographics, location of treatment, date of treatment, facilitator, facility, and accompanying person.

Author	First post	Gender	Age of patient when post was made (years)	City of procedure	Treatment date	Facilitator	Facility	Accompanying person
017	November 3, 2016	Male	58	Chennai	October 15	N/A[a]	Asian Joint Rebuilding Institute in Chennai	Wife
018	July 28, 2010	Female	N/A	N/A	N/A	N/A	N/A	N/A
019	September 13, 2013	N/A	N/A	N/A	N/A	N/A	N/A	N/A
020	January 7, 2014	N/A	N/A	N/A	N/A	N/A	N/A	N/A
021	August 4, 2009	Female	35-49	Margao	November 8	N/A	MY EYE	N/A
022	June 19, 2014	N/A	N/A	N/A	June 13	N/A	N/A	N/A
023	June 25, 2010	N/A	N/A	N/A	N/A	N/A	N/A	Other
024	July 3, 2016	Female	70	Kolkata	2014	N/A	N/A	Family
025	May 3, 2016	Male	N/A	Chennai	May 14	N/A	Asian Joint Rebuilding Institute in Chennai	Wife
026	September 27, 2015	Male	35-49	New Delhi	September 15	N/A	ReLEX Smile Lasik	N/A
027	November 24, 2012	Male	35-49	Mumbai	2012	N/A	Hinduja Hospital	N/A
028	January 23, 2012	Male	20-29	N/A	2010	N/A	N/A	N/A
029	January 14, 2010	Male	N/A	Mumbai	N/A	N/A	Saifee Hospital	N/A
030	August 8, 2015	N/A	N/A	Palakkad	2015	N/A	N/A	N/A
031	September 25, 2007	Male	40	Bangalore	N/A	N/A	N/A	N/A
032	March 24, 2017	Female	N/A	Mumbai	2008	N/A	N/A	Father
033	March 12, 2005	Male	N/A	Thiruvananthapuram, New Delhi	N/A	N/A	N/A	N/A
034	May 26, 2005	Male	35	Panaji	N/A	N/A	N/A	Wife

[a]N/A: not applicable.

Table 5. Health travelers' date of initial narrative, demographics, location of treatment, date of treatment, facilitator, facility, and accompanying person.

Author	First post	Gender	Age of patient when post was made (years)	City of procedure	Treatment date	Facilitator	Facility	Accompanying person
035	September 30, 2009	N/A[a]	N/A	N/A	N/A	N/A	N/A	Family
036	February 23, 2010	Male	47	New Delhi	April 16, 2010—gallbladder surgery; April 13, 2010—dental; April 14, 2010—physical exam; April 23, 2010—dental; April 2010—colonoscopy	N/A	Apollo Hospital Escorts Heart Institute	N/A
037	August 12, 2015	Female	49	N/A	N/A	N/A	N/A	N/A
038	April 20, 2016	Male	N/A	N/A	April 16	Indian-MedTrip Health Care Consultants	N/A	Sister
039	July 5, 2014	N/A	N/A	Mumbai	June 25, 2014	N/A	Mangal Anand Center	N/A
040	September 4, 2013	Female	50-64	Goa	February 1	N/A	N/A	Husband
041	February 24, 2016	Male	N/A	Pune	February 2, 2016	N/A	N/A	N/A
042	December 25, 2013	Male	27	N/A	2013	N/A	N/A	N/A

[a]N/A: not applicable.

Table 6. Health travelers' date of initial narrative, demographics, location of treatment, date of treatment, facilitator, facility, and accompanying person.

Author	First post	Gender	Age of patient when post was made (years)	City of procedure	Treatment date	Facilitator	Facility	Accompanying person
043	July 21, 2015	N/A[a]	N/A	N/A	N/A	Placidway	N/A	Spouse
044	April 17, 2014	Female	N/A	Bangalore	N/A	N/A	N/A	Daughter
045	November 21, 2013	Female	N/A	Mumbai	N/A	N/A	Rotunda—The Center for Human Reproduction	Husband
046	December 9, 2015	Female	N/A	Kartikulam	December 15	N/A	Ayurveda Yoga Villa	N/A
047	July 29, 2015	Female	41	Coimbatore	July 15	N/A	Vaidyagrama	N/A
048	March 19, 2012	Female	N/A	Coonoor, India	January 12	N/A	Ayurveda Yoga Retreat	N/A
049	September 12, 2012	N/A	N/A	N/A	N/A	N/A	N/A	N/A
050	December 4, 2006	Female	45	N/A	N/A	N/A	N/A	N/A
051	March 25, 2010	Female	N/A	Puttaparthi	N/A	N/A	N/A	Other
052	September 26, 2010	Male	N/A	Una	N/A	N/A	N/A	N/A
053	July 1, 2015	N/A	N/A	Mysore	N/A	N/A	Columbia Asia Hospital	N/A

[a]N/A: not applicable.

Reasons for Medical Tourism in India

The medical tourism drivers were patients attempting to receive the best care for a specific diagnosis, travelers trying to overcome home country barriers (ie, costs or unavailability), and patients seeking alternative treatments for a condition; there were a few who were just curious about the overall medical tourism experience and options. One of the more polarizing direct quotes was "No, I can't get the same treatment in America, that's why I'm here." Another stated "I have private health insurance, but I wasn't happy with the experience (or lack of) surgeons have here in carrying out the procedure."

It appeared that patients became open to the idea of medical tourism while researching one of two things. They were either looking for a provider who specialized in treating a certain diagnosis or researching procedures to treat certain conditions. A patient exploring Lasik options stated:

> *At 64 I decided that I had enough of broken glasses reading in bed. I went for a consult with a very reputable eye laser company here in Toronto. After all the tests, (no blood work) they can only restore my far vision at the cost of $6000+. I did some research and found a doctor in India…made an appointment through their site.*

In fact, there have been studies that confirm various health clinics will use the internet to advertise their services globally and that they not always clearly depict the risks of certain treatments [24].

Patient Satisfaction

The number of satisfied health travelers greatly outnumbers those who had complaints or expressed unhappiness. For example, some narrators made comments such as "happy with dental implants," "normal life regained," or just "happy with results." General problem resolution comments were the most abundant indicator that patients were satisfied with the care they received. With Ayurveda treatments, specifically, there were comments such as "found a love for yoga" or "found true self" that were not initially expected outcomes by the narrator. One Ayurveda patient said they found an "unexpected healing." A narrative, in which a patient received Lasik eye surgery, stated:

> *My sight is amazing and having had a session with Dr. Chandrakant Gaonker again this morning, he says it will get better each day.*

Another critical point surrounds the topic of desired repeat travel to receive care again. The repeat travel was not because the initial visit required any type of follow-up care, but it was because the patient was so happy with the experience, they wanted to go back to get another procedure. There was a narrator who mentioned they would only be getting their dental work completed in India moving forward. In addition, there was a narrator who considered permanently moving to India because they felt the care was highly superior to that offered in their home country for supporting specific chronic conditions.

When patients expressed negative feedback on the outcome and results, the negativity fit in the following primary categories: general bad experience, bad appearance, unresolved problem,

or issue worsened. Dental procedures most frequently garnered an unsatisfactory comment about the results, especially when relating to a narrator's appearance after care was given. After receiving a dental bridge, a patient said:

> *The curve of my finished bridge was way out of sync with my upper gum line, and the inner gum line sat too low from my gum, causing every bite I took to clog up in that space.*

One patient with Crohn disease did not mention that their condition worsened, or any corrections had to be made, but that the Ayurveda treatment had no impact:

> *Despite the doctor's assurances, it did not help much, I still had the same intolerances and same degree of inflammation.*

Perception Influencers

The influences that shaped the narrators' Indian medical tourism care experience included India's general environment. This was the noise, smells, and basic laws. One patient summarized his feelings by saying:

> *The problem is that I don't know the culture, don't speak the local dialect, have a hard time understanding the English of the average Indian, am a little intimidated by the system, and feel somewhat alone and isolated.*

Cost was another perception influencer; the costs mentioned were everything from the lodging to receiving care. Travel logistics (such as the passport and visa process), quality of care, customer service, facility condition, and follow-up care were the other main contributors that shaped narrators' opinions. For example, regarding costs, a notable trend among the narratives was that they described the payment process. Prepayment was one of the more frequent payment processes mentioned. Throughout all the narratives, there were 6 references to a prepayment requirement (before services were rendered). There was one narrator who mentioned they not only had to pay in advance, but they were required to pay cash:

> *He insists on only cash, so I handed over Rs. 98,500 (remember Rs. 1500 was paid already for the tests). I felt like I was doing some sort of underhand deal, however I fully trusted Dr. Sandip.*

Discussion

Principal Findings

Almost all the narratives came from patients seeking resolution of a health issue. There were no patients who went to India merely for exploratory purposes. Patients either felt they needed to fix their appearance (ie, cosmetic) or had a serious medical diagnosis such as a hernia. The patients found India by looking for not only the best care but also the most affordable resolution, which is consistent with the findings of the study by Ozan-Rafferty et al [9].

The general theme of all the narratives reviewed was positive. It was rare for a patient to indicate that their travel to India for care was negative. Many of the narratives focused on how affordable the care was when compared with that in the patients' home countries. An author even mentioned that he tried to give a financial tip to a provider because the care was so good and affordable. The other positive influencers were the narrators' good interactions with the health care providers in India, the modern facilities, and the resolution of their health issues. Many narrators mentioned that they would be a medical tourist in India again and encouraged others to travel to India for medical care. With that said, it is still important to state that there could be far more negative experiences, but the patients did not necessarily document their experiences on the web.

Reasons

The narrators said that a barrier prevented them from resolving their health issues in their home country and drove them to India. Some people sought procedures such as Ayurveda treatment, Desarda hernia repair, or limb lengthening that are simply not offered in all parts of the world. Some patients could not afford or did not have the type of insurance necessary to get the care they needed. A good example is the patients who went to India for dental care because they did not have dental insurance in their home countries. In fact, dental was the most common type of care the patients in this study received. There was an author who said they received a root canal at no charge.

The blog comments and discussion threads included detailed conversations between former and possible future patients just talking about medical tourism in general. It was obvious that some participants provided more insights to the readers in a more private format because they would reference a private message exchange with another user. Sometimes experiences were reconfirmed when other medical tourists commented that they had gone to India for care and had a similar story (ie, same doctor or facility).

Satisfaction Levels

As mentioned, many of the authors said their primary issue was resolved for an affordable price, and the authors encouraged others to consider medical tourism when looking to have their health issues resolved. As this study is replicated, the number of primarily negative-toned narratives is critical, similar to those highlighted in the next section; it will help to answer the question, if providers of medical tourism are improving (based on feedback) or ignoring opportunities to get better. In addition, patients' opinions, which were once positive, could change if there are any delayed side effects.

Negative Viewpoints

All of the negative comments tied back to the health issue not being fixed to the patient's satisfaction. This section will focus on two of the narrators who were the most dissatisfied. A dental patient's condition worsened as a result of bad implants, and the author realized it would be difficult for her to file a lawsuit because she voluntarily went out of her home country to get care. Another narrator said that providers spoke to each other in an unidentified language even though they all spoke English, and this made her feel excluded and uncomfortable. That same patient said she believed some of the utensils used during preoperational procedures were not clean.

Interactions With Hospitals and Providers

The number of patients and their companions who reported good encounters with the providers in India greatly outnumbered those who experienced bad encounters. The positive comments included, but were not limited to, providers being well dressed, providers having extensive experience, friendly doctors, and some patients being greeted with flowers upon arrival. There were more than a few times that a provider had good web-based reviews, which initially led the patient to the provider. In addition, patients really seemed to be comforted if the provider had experience or training in the United States, and a large number of the patients indicated the providers either spoke English or had a translator readily available.

When negative comments were made specifically related to provider and patient interaction, they seemed to be aimed at inaccurate web-based advertisements of the doctor's capabilities and specialties and general poor communication. For example, one patient landed in India only to find out their appointment had been canceled, and another narrative described a provider as being money hungry.

Challenges to Medical Tourism in India

Some of the main challenges to medical tourism in India are poor visitor experiences expressed in negative web-based reviews, travel logistics (ie, long flights, having to get passports and travel visas), lack of capital compared with other countries marketing medical tourism, and improved health care infrastructure across the globe [13]. Regarding negative web-based reviews, many of the narratives indicated that the patient completed web-based research before deciding to go to India for care, and it was not unusual for patients to remark that it was difficult to identify trustworthy information or find information about medical tourism in India. There were previous medical tourism patients who had posted negative feedback about doctors on web-based forums or blogs and urged others to avoid medical tourism. In the narratives used for this study, a dental patient said the dentist they saw in India had persuasive marketing but was not successful in resolving any of the patient's primary diagnosis. The patient said the provider deleted any bad reviews users posted on Facebook.

Although India does have a medical tourism–focused visa, the passport and visa process creates another hurdle for patients who want to go to India for care. Once the appropriate passports and visas are obtained, there can be the issue of discomfort owing to a long flight.

India is not the only country known as a medical tourism destination, and some countries are working hard to not lose patients to medical tourism. Within this study, there was a narrative in which a dental patient wrote:

> *I researched nearly every clinic in New Delhi and trust me—none of them are as well trained and professional as in the U.S. YOU GET WHAT YOU PAY FOR!!! My suggestion is to try Buenos Aires, Argentina.*

Brazil, Turkey, Mexico, Costa Rica, and Thailand are just a few of the other countries promoting medical tourism [25]. These options could be more attractive to a potential medical tourism patient if they are closer, cheaper, or do not require a travel visa.

Opportunities for Medical Tourism Improvement in India

After reviewing the narratives, three opportunities for the improvement of medical tourism in India are apparent. The first is international accreditation. Several patients did not have a process to ensure they were receiving quality service other than "everything looked up to date and clean." Although there are hospital-accrediting organizations and processes in India, the certification and approval details need to become more prominent and widely advertised. Adding to that, organizations such as Healthgrades and RateMD could expand to emphasize provider reviews specifically tailored for medical tourists.

There is an opportunity for Indian providers to partner with providers across the globe to assist with patient preparation and follow-up care. Many narratives mentioned that there was additional time spent in India to prepare for their procedure (such as x-rays, bloodwork) or get the appropriate follow-up care. Patients mentioned that they had to do substantial self-care once they went back to their home country, such as changing dressings or preparing food in a certain manner. If an Indian doctor partnered with providers in other countries, this could reassure patients that their needs could easily be met before and after their care was completed in India, eliminating the cumbersome follow-up care process.

The third item is the extremely limited availability of medical tourism insurance that covers the costs of procedures completed overseas. Although the narratives did mention huge cost savings when doing an "apples to apples" comparison (price of procedure in home country vs price in India), the patients were paying out of pocket for travel, lodging, and care. There is a huge opportunity for India to introduce insurance that would help alleviate some of the current medical tourism out-of-pocket costs.

Limitations

There are no official statistics or databases regarding how many people travel internationally to obtain health care [26]. So, there was no pool of known health travelers that could have been contacted to gather general additional insight or expedite the narrative search process.

Greenhalgh [27] pointed out that a primary limitation of narrative research is that narratives are open to multiple interpretations. With that, Bhattacharya [28] added that researchers need to be mindful of their own subjectivities; they should be aware of how their personal experiences creep into the research process and influence the way in which the researcher understands the narrations.

Following the method established in the study by Ozan-Rafferty et al [9], there were already some built-in limitations. These limitations included the chance for fictitious and biased data to be present in the narratives, and the researcher had limited methods to validate self-reported information [29]. As described in the Methods section, narratives were not used if they were posted for promotional purposes. This would include

testimonials posted on the actual webpages of providers, websites that strictly focused on consumers rating their experiences, or any post that flat out said something such as "the provider asked me to post this..." Anonymity releases authors from any consequences of their posts or behavior, but at the same time, anonymity may encourage people to share more on the web and discuss sensitive topics and issues [30]. The sample did not include individuals who lacked access to a computer or some type of internet-connected device (to post their narrative on the World Wide Web) and thus may not be representative of the general population [31]. Of the 53 narratives, only 6 were blogs. As these were not live interactions (ie, real-time interviews or conversations), the researchers were unable to clarify feedback provided by the health travelers [9].

Since Ozan-Rafferty et al [9] completed their original study related to this topic, Google has disabled the search filter that allowed users to limit search results strictly to blogs. The feature was retired in 2014 [32].

Comparison With Prior Work

It is notable that a portion of this study's results align with some of the results in the study by Ozan-Rafferty et al [9], which focuses on a completely different country. For example, Europe and North America were the top home origination countries for medical tourists. Most of the patients were content with the care they received while being abroad, and the health travelers thought fondly of the providers they saw in the host countries. One of the negative commonalities that stood out was the bad traffic described in both Turkey and India. The results of studies that follow this same method but apply it to different countries can be combined to make more precise conclusions about medical tourism, overall.

There was a contrast in the popular procedures from each study. In the study by Ozan-Rafferty et al [9], the top procedure that medical tourists availed was hair transplants. However, only 1 patient mentioned getting a hair transplant when India was studied. Ayurveda was the second most mentioned procedure in this study, but it was not mentioned at all in the analysis by Ozan-Rafferty et al [9].

Conclusions

This is not the first study to focus on medical tourism or medical tourism in India, specifically. As mentioned, it was actually birthed from the recommendations of a prior study, but it is unique in how and when the research occurred and the method that was used to study medical tourists who traveled to India for care. Although keeping technology and the medical industry constantly evolving, the core of this study's findings can be beneficial to health care providers, patients, and governments. The method can be applied to research medical tourism in other countries or even to gather consumer feedback on topics beyond health care. About 25 years ago, this study would not have been possible. The advancement of internet search engines and increased socializing on the internet has greatly helped to evolve research. Now, there are actual E-Hospitals. Hong [33] wrote about the Chinese American Physician E-Hospital's October 2015 grand opening. All physicians affiliated with this E-Hospital are bilingual Chinese American physicians who are board certified in the United States, and the website is designed specifically for the Chinese population (webpage only available in Chinese). The services provided include international transfer in the United States, saving the patients the hassle of identifying and connecting with an appropriate health service provider and also minimizing language and cultural barriers [33].

The limitations previously stated must still be considered, specifically considering the fact that there is no guaranteed way to assure all the narratives are truthful, and there is no solid estimate on how many medical tourists are actually traveling to India. However, providers hosting medical tourists in India can use the negative feedback to make corrections in the services they provide. As many patients noted disorganization was a sore point in their experience, providers can make sure to have organized lines when patients are checking in or out of the facilities. Providers who are losing patients to medical tourism can refer to this study and make service adjustments so that fewer patients are swayed to become medical tourists. For example, more providers might start practicing forms of Ayurveda, which is currently not provided in many countries outside of India. Many patients in the study specifically went to India to receive Ayurvedic care. Some narratives mentioned extending the length of trips so patients could receive appropriate follow-up care and some mentioned that follow-up care was not possible in their home countries. Providers can seek ways to enhance the follow-up care process by possibly setting up satellite offices in partnership with providers in other countries.

These results open up more options for patients. There are positives (ie, very affordable) and negatives (ie, bad traffic) about medical tourism in India. In addition, prospective patients can use altered and less-intensive versions of this study's method to research how they can resolve their specific health issues as medical tourists in India or in other countries. In addition, patients who have previously been medical tourists and review this project can add to the details provided by authors of the narratives. More participation will help increase the accuracy of future work.

Acknowledgments

This research was conducted as part of the first author's (JB) requirements for completing a Doctorate of Health Administration at Central Michigan University. The author would like to thank the Dissertation Committee Chair, JJ; Dissertation Committee Member, MO; Dissertation Committee Member, MS; and SB for coding review.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Steps in the Narrative Process.
[PNG File , 265 KB - jmir_v22i5e15665_app1.PNG]

Multimedia Appendix 2
Theme frequency count table.
[PNG File , 57 KB - jmir_v22i5e15665_app2.PNG]

References

1. Fetscherin M, Stephano R. The medical tourism index: scale development and validation. Tour Manag 2016 Feb;52:539-556. [doi: 10.1016/j.tourman.2015.08.010]
2. Connell J. Contemporary medical tourism: conceptualisation, culture and commodification. Tour Manag 2013;34:1-13. [doi: 10.1016/j.tourman.2012.05.009]
3. Mutalib NS, Ming LC, Yee SM, Wong PL, Soh YC. Medical tourism: ethics, risks and benefits. Indian J Pharm Educ Res 2016;50(2):261-270 [FREE Full text] [doi: 10.5530/ijper.50.2.6]
4. Lunt N, Horsfall D, Hanefeld J. Medical tourism: a snapshot of evidence on treatment abroad. Maturitas 2016 Jun;88:37-44. [doi: 10.1016/j.maturitas.2016.03.001] [Medline: 27105695]
5. Hegde R. Quora. Why is Medical Education So Costly in India, and Why Recently Has the Cost of It Increased So Dramatically? URL: https://www.quora.com/Why-is-medical-education-so-costly-in-India-and-why-recently-has-the-cost-of-it-increased-so-dramatically [accessed 2018-08-31] [WebCite Cache ID 725bsvqcJ]
6. Alsharif MJ, Labonté R, Lu Z. Patients beyond borders: a study of medical tourists in four countries. Glob Soc Policy 2010;10(3):315-335. [doi: 10.1177/1468018110380003]
7. Frost JH, Massagli MP. Social uses of personal health information within PatientsLikeMe, an online patient community: what can happen when patients have access to one another's data. J Med Internet Res 2008 May 27;10(3):e15 [FREE Full text] [doi: 10.2196/jmir.1053] [Medline: 18504244]
8. Abubakar AM, Ilkan M. Impact of online WOM on destination trust and intention to travel: a medical tourism perspective. J Dest Mark Manag 2016;5(3):192-201. [doi: 10.1016/j.jdmm.2015.12.005]
9. Ozan-Rafferty ME, Johnson JA, Shah GH, Kursun A. In the words of the medical tourist: an analysis of Internet narratives by health travelers to Turkey. J Med Internet Res 2014 Feb 6;16(2):e43 [FREE Full text] [doi: 10.2196/jmir.2694] [Medline: 24513565]
10. Johnson JA, Stoskopf C, Shi L. Comparative Health Systems: A Global Perspective. Burlington, MA: Jones & Bartlett Learning; 2010.
11. Gupta V, Das P. Medical tourism in India. Clin Lab Med 2012 Jun;32(2):321-325. [doi: 10.1016/j.cll.2012.04.007] [Medline: 22727009]
12. Ministry of Health and Family Welfare, Government of India. Ministry of Health and Family Welfare. National Health Policy 2017 URL: https://mohfw.gov.in/sites/default/files/9147562941489753121.pdf [accessed 2017-07-19]
13. Sharma A. Medical tourism: emerging challenges and future prospects. Int J Bus Manag Invent 2013;2(1):21-29 [FREE Full text]
14. Lunt N, Carrera P. Medical tourism: assessing the evidence on treatment abroad. Maturitas 2010 May;66(1):27-32. [doi: 10.1016/j.maturitas.2010.01.017] [Medline: 20185254]
15. Saltsman W, Lett J. You had it done where? the rise of medical tourism. Caring Age 2017;18(4):12. [doi: 10.1016/j.carage.2017.03.008]
16. Bernasek A. Newsweek. 2015 Jan 6. Two Numbers: Sun, Sea and Stitches URL: https://www.newsweek.com/2015/01/16/two-numbers-sun-sea-and-stitches-297090.html [accessed 2018-08-19] [WebCite Cache ID 71nXqqQux]
17. Gupta M, Rajachar V, Prabha C. Medical tourism: a new growth factor for Indian healthcare industry. Int J Res Med Sci 2015;3(9):2161-2163. [doi: 10.18203/2320-6012.ijrms20150597]
18. Creswell JW. Qualitative Inquiry and Research Design: Choosing Among Five Approaches. Newbury Park, CA: Sage Publications, Inc; 2013.
19. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Adm Policy Ment Health 2015 Sep;42(5):533-544 [FREE Full text] [doi: 10.1007/s10488-013-0528-y] [Medline: 24193818]
20. Cooper DR, Schindler P, Blumberg B. Business Research Methods. London: McGraw-Hill Education; 2011.
21. O'Brien MR, Clark D. Unsolicited written narratives as a methodological genre in terminal illness: challenges and limitations. Qual Health Res 2012 Feb;22(2):274-284. [doi: 10.1177/1049732311420737] [Medline: 21876209]
22. O'Leary DE. Blog mining-review and extensions: 'from each according to his opinion'. Decis Support Syst 2011;51(4):821-830. [doi: 10.1016/j.dss.2011.01.016]
23. Neuman WL. Social Research Methods: Qualitative and Quantitative Approaches. Boston: Pearson; 2009.

24. Kashihara H, Nakayama T, Hatta T, Takahashi N, Fujita M. Evaluating the quality of website information of private-practice clinics offering cell therapies in Japan. Interact J Med Res 2016 May 24;5(2):e15 [FREE Full text] [doi: 10.2196/ijmr.5479] [Medline: 27222494]
25. Stephano RM. Medical Tourism Magazine.: Medical Tourism Magazine Top 10 Medical Tourism Destinations in the World URL: https://www.magazine.medicaltourism.com/article/top-10-medical-tourism-destinations-world [accessed 2018-09-17] [WebCite Cache ID 71nXkijZf]
26. Pollard K. Pollard K.: International Medical Travel Journal; 2017 Jan 26. India Medical Tourism Market: Worth $8 Billion? The Reality URL: https://www.imtj.com/blog/india-medical-tourism-market-worth-8-billion-reality/ [accessed 2018-08-19] [WebCite Cache ID 71nXC42xB]
27. Greenhalgh T. Cultural Contexts of Health: The Use of Narrative Research in the Health Sector. Geneva: World Health Organization; 2016.
28. Bhattacharya A. The many ways of knowing: embracing multiplicity in narrative research. Qual Soc Work 2016;15(5-6):705-714. [doi: 10.1177/1473325016652683]
29. Jones M, Alony I. Blogs – the new source of data analysis. Issue Inform Sci Inf Technol 2008;5:433-446 [FREE Full text] [doi: 10.28945/1019]
30. Pfeil U, Zaphiris P. Applying qualitative content analysis to study online support communities. Univ Access Inf Soc 2009;9(1):1-16. [doi: 10.1007/s10209-009-0154-3]
31. Seale C, Charteris-Black J, MacFarlane A, McPherson A. Interviews and internet forums: a comparison of two sources of qualitative data. Qual Health Res 2010 May;20(5):595-606. [doi: 10.1177/1049732309354094] [Medline: 20008955]
32. Brinkmann M. gHacks Technology News. 2014 Jan 23. Here is How You Search Discussions, Blogs, Places on Google Now URL: https://www.ghacks.net/2014/01/23/search-discussions-blogs-places-google-now/ [accessed 2018-08-19] [WebCite Cache ID 71nYxoI6V]
33. Hong YA. Medical tourism and telemedicine: A new frontier of an old business. J Med Internet Res 2016 May 23;18(5):e115 [FREE Full text] [doi: 10.2196/jmir.5432] [Medline: 27215230]

Edited by G Eysenbach; submitted 27.07.19; peer-reviewed by G Shah, SGS Shah; comments to author 04.10.19; revised version received 28.11.19; accepted 24.01.20; published 14.05.20.

Please cite as:
Brown J, Johnson J, Ozan-Rafferty ME, Sharma M, Barbera S
Internet Narratives Focused on Health Travelers' Experiences in India: Qualitative Analysis
J Med Internet Res 2020;22(5):e15665
URL: https://www.jmir.org/2020/5/e15665
doi:10.2196/15665
PMID:32406860

Original Paper

Paper Versus Digital Data Collection Methods for Road Safety Observations: Comparative Efficiency Analysis of Cost, Timeliness, Reliability, and Results

Niloufer Taber[1], MPA, MPH; Amber Mehmood[1], MBBS, MPH, FCPS; Perumal Vedagiri[2], ME, PhD; Shivam Gupta[1], MPH, MD, PhD; Rachel Pinto[1], BDS, MPH; Abdulgafoor M Bachani[1], MPH, PhD

[1]International Injury Research Unit, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[2]Department of Civil Engineering, Transportation Systems Engineering, Indian Institute of Technology Bombay, Mumbai, India

Corresponding Author:
Niloufer Taber, MPA, MPH
International Injury Research Unit
Johns Hopkins Bloomberg School of Public Health
615 N Wolfe St
Baltimore, MD, 21205
United States
Phone: 1 410 955 3934
Email: ntaber1@jhu.edu

Abstract

Background: Roadside observational studies play a fundamental role in designing evidence-informed strategies to address the pressing global health problem of road traffic injuries. Paper-based data collection has been the standard method for such studies, although digital methods are gaining popularity in all types of primary data collection.

Objective: This study aims to understand the reliability, productivity, and efficiency of paper vs digital data collection based on three different road user behaviors: helmet use, seatbelt use, and speeding. It also aims to understand the cost and time efficiency of each method and to evaluate potential trade-offs among reliability, productivity, and efficiency.

Methods: A total of 150 observational sessions were conducted simultaneously for each risk factor in Mumbai, India, across two rounds of data collection. We matched the simultaneous digital and paper observation periods by date, time, and location, and compared the reliability by subgroups and the productivity using Pearson correlations (r). We also conducted logistic regressions separately by method to understand how similar results of inferential analyses would be. The time to complete an observation and the time to obtain a complete dataset were also compared, as were the total costs in US dollars for fieldwork, data entry, management, and cleaning.

Results: Productivity was higher in paper than digital methods in each round for each risk factor. However, the sample sizes across both methods provided a precision of 0.7 percentage points or smaller. The gap between digital and paper data collection productivity narrowed across rounds, with correlations improving from r=0.27-0.49 to 0.89-0.96. Reliability in risk factor proportions was between 0.61 and 0.99, improving between the two rounds for each risk factor. The results of the logistic regressions were also largely comparable between the two methods. Differences in regression results were largely attributable to small sample sizes in some variable levels or random error in variables where the prevalence of the outcome was similar among variable levels. Although data collectors were able to complete an observation using paper more quickly, the digital dataset was available approximately 9 days sooner. Although fixed costs were higher for digital data collection, variable costs were much lower, resulting in a 7.73% (US $3011/38,947) lower overall cost.

Conclusions: Our study did not face trade-offs among time efficiency, cost efficiency, statistical reliability, and descriptive comparability when deciding between digital and paper, as digital data collection proved equivalent or superior on these domains in the context of our project. As trade-offs among cost, timeliness, and comparability—and the relative importance of each—could be unique to every data collection project, researchers should carefully consider the questionnaire complexity, target sample size, implementation plan, cost and logistical constraints, and geographical contexts when making the decision between digital and paper.

(J Med Internet Res 2020;22(5):e17129) doi:10.2196/17129

KEYWORDS

information technology; public health informatics; mHealth; risk factors; population surveillance; traffic accidents; data collection

Introduction

Background

Road traffic injuries (RTIs) are a major global public health problem. With over 1.35 million deaths and 20-50 million nonfatal injuries estimated each year, RTIs impact all age groups and populations of all socioeconomic backgrounds [1]. However, low- and middle-income countries (LMICs), with rapid motorization in an unsafe road environment, bear a disproportionate share of deaths and disability [1]. In addition to safe vehicles, road infrastructure, and road safety management capacity, enhancing safe road user behavior plays an important part in preventing crashes and injuries [2]. The development and successful implementation of comprehensive programs to positively impact speeding, helmet, and seatbelt wearing require rigorously monitoring the prevalence of behavioral risk factors. One such example is the Bloomberg Initiative for Global Road Safety (BIGRS), a multisectoral program that unites a consortium of partners that work together to reduce the burden of RTIs in 10 cities in LMICs [3]. Roadside observational studies have played a fundamental role in designing evidence-informed strategies, including media campaigns, enforcement, and environmental modifications to enhance road safety.

Paper-based data collection has been the standard method for primary observational studies, but due to the possibility of human errors, storage costs, and time and labor required for double data entry, digital data collection methods are gaining popularity. Two concerns frequently influence the decision to switch from paper to digital methods of data collection. First, whether digital data collection is as productive and reliable as paper data collection in a dynamic roadside environment, and second, whether digital data collection is at least as efficient as paper data collection. Recent observational studies of road safety risk factors in three different countries, using both approaches have consistently demonstrated reliable results between the two methods [4]. In LMICs, with limited resources and relatively less expensive labor costs, paper data collection has been traditionally deemed as feasible and potentially more affordable. However, the efficiency of digital data collection vs traditional paper-based approaches has not been empirically assessed in these settings.

Efficiency is a rate measure against cost, time, or accuracy, and is used to assess whether the desired output can be produced in less time, using fewer resources, or with fewer errors. Efficiency is contextual, as cost and time efficiency is tied to the availability of infrastructure or other logistical requirements, as well as human and material resources. For a researcher deciding between using digital or paper data collection, provided there is a reasonable level of reliability and comparable productivity between the two methods, the decision may be based on factors related to efficiency.

Objectives

We conducted an observational study in Mumbai, India, to understand the reliability, productivity, and efficiency of paper vs digital data collection based on three different road user behaviors: helmet use, seatbelt and child restraint use, and speeding. We also conducted a cost and time comparison to understand the relative efficiency of paper vs digital data collection across successive rounds, and to evaluate the potential trade-offs among different dimensions of efficiency (Table 1).

Table 1. Three areas of comparison between digital and paper data collection: productivity, reliability, and efficiency.

Dimensions of each area of comparison	Methods of measurement
Productivity	
Volume	• Number of observations per observation session
Precision	• Margin of error for estimation of proportions • Akaike Information Criteria in regression analysis
Reliability	
Statistical reliability	• Proportion of risk factor by date, time, and location, as well as vehicle, occupant, and environmental characteristics
Comparability of results	• Adjusted odds ratios for vehicle, occupant, and environmental risk factors
Efficiency	
Cost	• Per survey • Per dataset to achieve a certain level of precision • Per labor-hour of time (may be differential by skill level and cost of labor)
Time	• Per survey • Per complete dataset (preparation, data collection, data entry and verification, data management and cleaning)

Methods

Setting

This comparative study was conducted in Mumbai. Mumbai is India's most populous city, as well as the country's financial center and commercial capital [5]. Mumbai city has a population of approximately 12.5 million, with another 20.6 million people in the metropolitan area [6]. Mumbai has a high literacy rate, close to 90%, with 15.8 million internet subscribers [7,8]. Mumbai ranks sixth among major cities in India for number of road traffic collisions, with over 82% of all crashes resulting in at least one injury and almost 15% of collisions resulting in at least one fatality [9]. The registered vehicle fleet was 2,571,000 in 2015 [10].

Data Collection

As part of the BIGRS project, the Johns Hopkins International Injury Research Unit partnered with the Indian Institute of Technology, Bombay to conduct semiannual observational studies to measure the prevalence of helmet use, seatbelt and child restraint use, and speeding at representative locations throughout the city. The details of the roadside data collection protocol are provided elsewhere [4]. To assess reliability between the two methods, paper and digital data collection were conducted simultaneously between January-March 2018, and again between July-September 2018. In each round, 75 of 150 observation sessions had simultaneous paper and digital data collection for each risk factor, which was used for our study analysis. Consistent procedures and definitions were maintained between methods, across observation sites, and across rounds to ensure comparability of results.

Observations on helmet and seatbelt use were conducted at intersections while vehicles were stationary to allow field teams to observe the use of safety equipment, whereas speeding observations were conducted at unobstructed stretches of road. Data were captured by an observer, who kept his or her attention on the flow of traffic and reported observations to a recorder, who wrote or entered into a tablet the road safety data. During the initial years of the project, the entire data collection was paper-based, while pilot testing of digital data collection in Mumbai started in 2018.

Each observation session began by capturing information about the site and observation session. These included (1) date, time of day, and location of each observation period; (2) names of the observer and recorder for that session; (3) the weather; (4) the volume of traffic over time; and (5) the presence of any law enforcement during the session. Paper-based data collection was handwritten into forms with a predesigned grid appropriate to each risk factor. Each row captured information on a single vehicle. For speeding observations, the columns captured road safety behavior for each vehicle, whereas for helmet and seatbelt use observations, occupant position, demographics, and road safety behavior were captured in a set of columns for each occupant in the vehicle.

The digital data collection tool was developed to mirror the paper-based tool, capturing all the same information, but in a series of screens through which the recorder would swipe to record information on each vehicle and vehicle occupant. For the BIGRS project, we used KoBoToolbox for digital data collection, an open-source mobile data collection platform [11]. The digitization process involved programming digital forms that were then downloaded onto the KoBoCollect mobile app. The content of both paper and digital questionnaires was the same; however, the digital form included mandatory fields, logic checks, and constrained text entry that prevented recorders from leaving fields blank or entering unreasonable or inconsistent responses. The team used Android tablets for data collection and uploaded the questionnaire forms to a secure cloud server at the end of each session.

Local data collectors were trained for two days on study protocols, paper data entry, the use of Android tablets and the KoBoCollect app, and how to upload data to the server. Supervisors and data managers were trained to monitor field site data collection and data aggregation on the server.

Statistical Analysis

Helmet use was categorized as correctly, incorrectly, or not used. Correct helmet use was defined as using a strapped, standard helmet, whereas incorrect helmet use involved either wearing a *cap* or *tropical* helmet, or using an unstrapped helmet. Neither digital or paper data collectors would be able to distinguish an imitation or substandard helmet from a genuine standard helmet, as that would require close inspection for a standards label. Although this may slightly inflate the proportion of the sample assessed to be wearing a helmet correctly, it would not do so differentially by data collection method, and so should not affect measures of comparison between the two methods. Correct seatbelt use was defined as the use of age-appropriate restraints: seatbelts for adults and child or booster seats for children under 12 years old. Incorrect and no restraint use were grouped due to the low occurrence of incorrect restraint use. Speeding was defined as any excess of the speed limit or was grouped categorically as not speeding, speeding up to 10 kilometers per hour (kph), and speeding over 10 kph above the posted limit.

For helmet use observations, the age variable captured whether the occupant was observed to be over or under the age of 18, whereas for seatbelt use observations, a greater level of detail was captured to assess whether children under 12 years were using age-appropriate child restraints. During the analysis, we did not distinguish children under 12 by gender, as observers were not able to reliably determine a child's gender through observation. Helmet use observations categorize motorcycle riders as drivers or passengers, whereas seatbelt use observations distinguished passengers further as sitting in the front or rear seats. Sex was captured as either male or female for both helmet and seatbelt use observations. Speeding observations did not capture any occupant characteristics, as data collectors cannot accurately observe inside vehicles in motion. For speeding and seatbelt use observations, the vehicle type and ownership were captured; vehicle types were either a sedan, sport utility vehicle, pickup or light truck, minivan, heavy truck, bus, or, for speeding observations only, motorcycle. Vehicles ownership was classified as private, government, commercial, taxi, or tourist vehicles, based on markings on the outside of the vehicle,

particular to the context. Vehicle type and ownership were not included in helmet use observations, as all vehicles were uniformly private motorcycles.

We matched simultaneous digital and paper observation periods by date, time, and location. We compared the productivity of the two methods by calculating the correlation in the number of observations made on vehicles (for speeding) or vehicle occupants (for helmet and seatbelt use) per 90-min observation session. The measure for reliability was the correlation in the prevalence of each risk factor among subgroups defined by vehicle occupant, vehicle, and environmental characteristics within matching sessions. Both reliability and productivity were assessed using Pearson correlations. Although Spearman rank correlations are more conventionally used to compare proportions, we were interested in how closely the proportions exactly matched each other, rather than whether the rank ordering of proportions was similar between digital and paper data collection methods. Pearson correlations are also appropriate for proportions when the proportions are not close to 0 or 1. We also performed a precision analysis, evaluating the margin of error available with current data and the sample size that would be required to estimate a proportion within 1 and 2 percentage points.

We also wished to understand whether the method of data collection would influence the results of descriptive or inferential analysis. To achieve this, we aggregated together all observation sessions by data collection method and round of observation (January-March 2018 and July-September 2018), creating four datasets. We conducted descriptive analysis separately within each dataset and compared the prevalence of various risk factors, overall and across subgroups. We also conducted multivariable logistic regression separately using each of the four datasets and compared the adjusted odds ratios.

Covariates included in the regressions included the occupant sex, age, and position in or on the vehicle, vehicle type, and ownership. Also included were the time of day, the location, and whether the date was a weekend or weekday. Finally, environmental factors included whether there was visible police presence, camera enforcement, or, for speeding observations, speed deterrents such as speed bumps, stop signs, or crosswalks. For analysis related to speeding, occupant characteristics were not included. As the location of the observation being naturally collinear with the presence of environmental speed deterrents, we included speed deterrents in our models and omitted locations, as the effect of environmental deterrents on speeding is of interest to road safety researchers.

Cost Estimates

Personnel costs included field data collectors (observers, recorders, and field supervisors) and data entry operators, and logistical costs included staff training and transportation to observation sites. Supplies and equipment included running costs for office supplies, space for paper data entry, data plans for tablets, with one-time costs for computers for data entry, tablets and power banks, and app development. We assumed that tablets and power banks would need to be replaced every 3 years. Digital data collection running costs also included annual server costs and cloud data backup. Costs were calculated using the 2018 midyear exchange rate between Indian rupees (INR) and US dollars. Costs were calculated for only the 75 observation sessions that were simultaneously collected using both paper and digital methods.

Time Estimates

The time to complete an observation on a vehicle (for speeding observations) or on the vehicle occupants (for helmet use and seatbelt use observations) was assessed by dividing the 90 minutes of each observation session by the volume of observations in that session, and then multiplying by 60 seconds for a minute to obtain the number of seconds needed to complete an observation. Finally, we took the average number of seconds to complete an observation for each risk factor in each round of data collection.

The time to obtain a complete dataset was measured as the number of days it took until the dataset was entered, cleaned, and ready for analysis, which included all work performed by the local partners onsite, including training, field work itself, data entry and verification of paper data, and both local and offsite data cleaning and management. As only half of the paper sessions were simultaneously collected digitally, time estimates for paper data entry and verification were appropriately divided in half.

Results

Productivity Comparison

Across each of the three risk factors, paper methods showed higher productivity than digital methods, in both the winter (January-March) and summer (July-September) data collection rounds (Figures 1-3). However, the gap between digital and paper data collection volumes narrowed between the winter and summer rounds, with the correlation in productivity by observation session increasing from r=0.27 to 0.96 among helmet use, 0.32 to 0.95 among seatbelt use, and 0.49 to 0.89 among speeding observations.

The results of our precision analysis showed that the existing sample sizes in digital and paper provided precision to within 0.7 percentage points or smaller (1.4% margin of error). Conversely, the difference between paper and digital levels of precision was less than 0.2 percentage points, regardless of risk factor or round of data collection (Table 2). When comparing results from the regression analysis, digital data collection had a lower Akaike Information Criteria (AIC) in all cases except in the summer speeding data collection. This was despite digital data collection methods having a smaller sample size and the same number of variables in the models (Multimedia Appendix 1).

Figure 1. Correct helmet use: correlation between digital and paper volumes of observations by round.

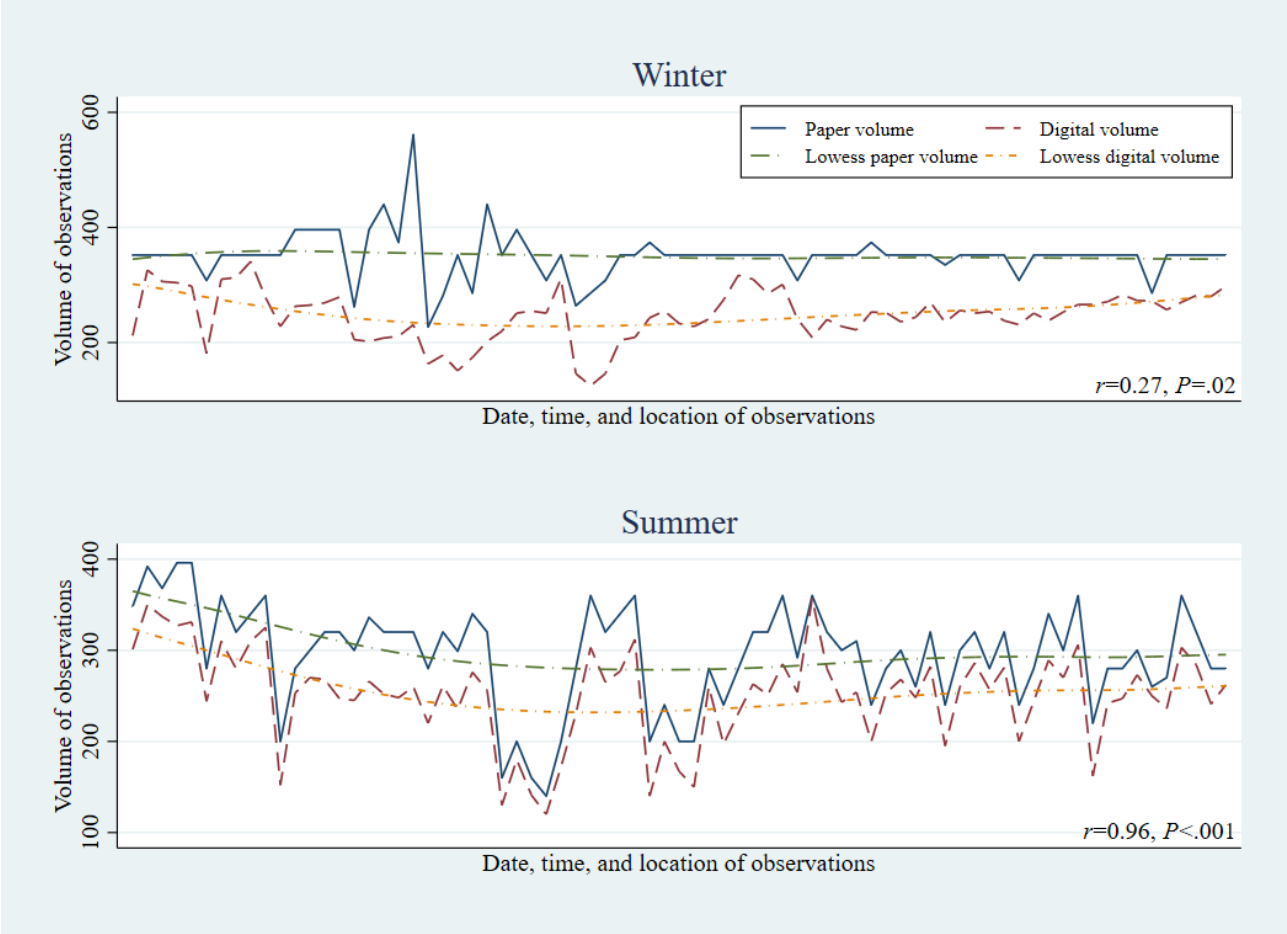


Figure 2. Seatbelt use: correlation between digital and paper volumes of observations by round.

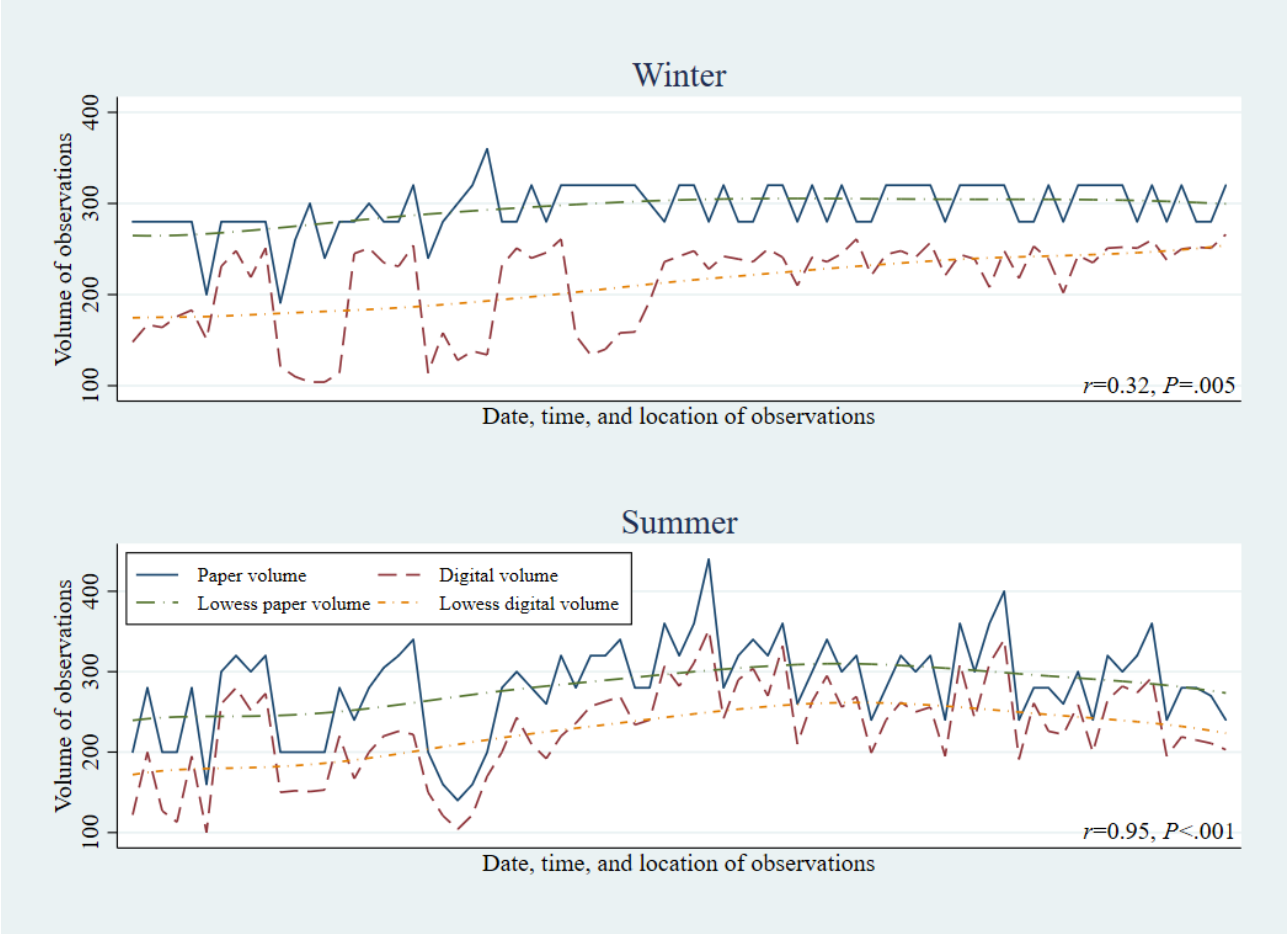


Figure 3. Speeding: correlation between digital and paper volumes of observations by round.

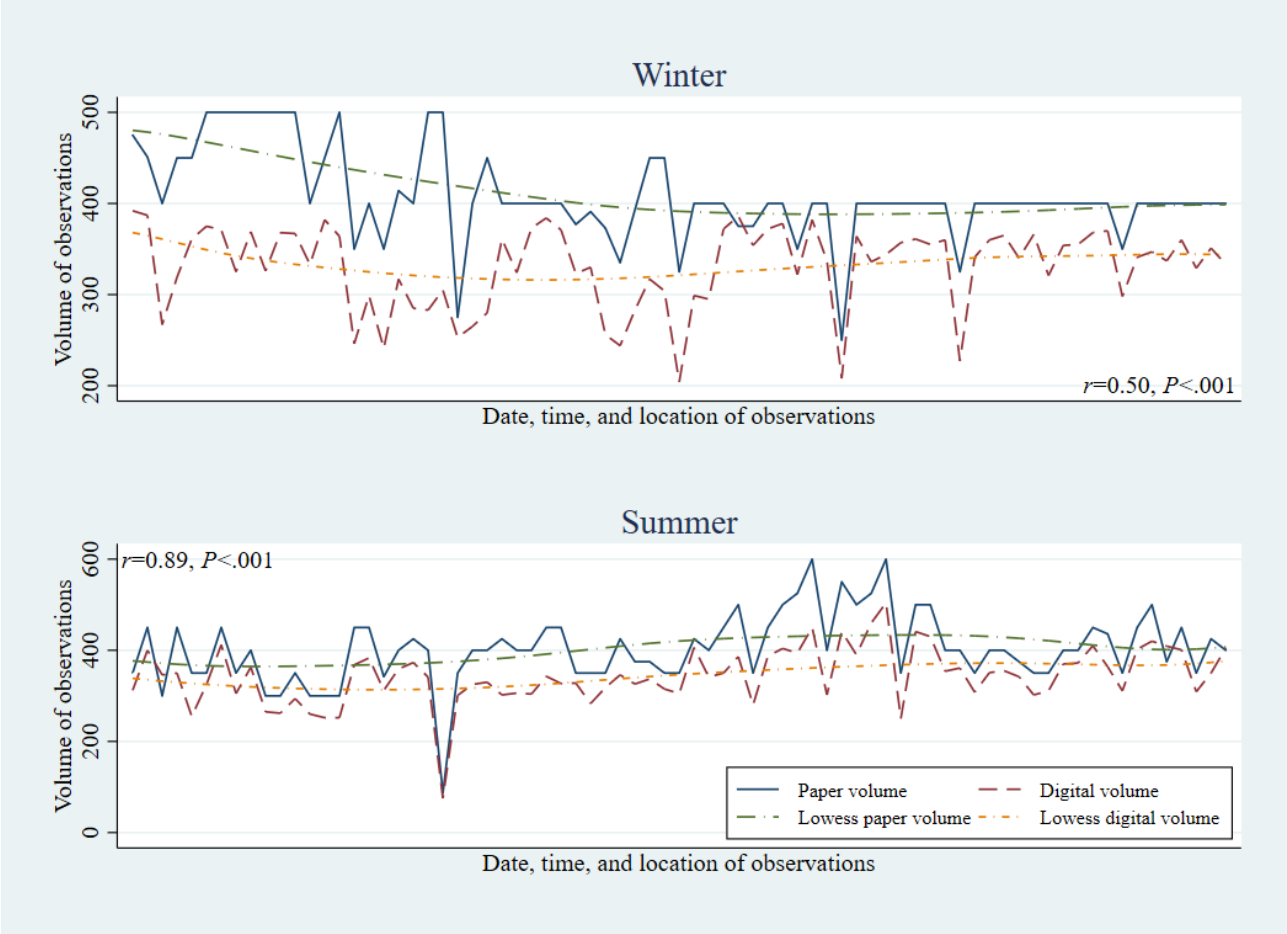


Table 2. Level of precision: current and needed sample sizes.

Risk factor and round of data collection	Paper sample size, n	Precision with existing paper sample, proportionate terms	Digital sample size, n	Precision with existing digital sample, proportionate terms	Required sample size for estimation within 1 percentage point (0.01), n	Required sample size for estimation within 0.5 percentage points (0.005), n
Helmet use						
Winter	34,309	0.006	24,283	0.007	9595	38,377
Summer	29,286	0.006	24,778	0.007	9585	38,338
Seatbelt Use						
Winter	36,573	0.006	25,452	0.007	9603	38,411
Summer	35,434	0.006	27,423	0.006	9604	38,414
Speeding data						
Winter	30,634	0.004	24,799	0.006	7827	31,307
Summer	30,190	0.004	25,783	0.006	8459	33,835

Statistical Reliability

Reliability in the prevalence of behavioral risk factors matched by date, time, location, and by characteristics of vehicles and vehicle occupants showed moderate to high levels of correlation in both rounds of data collection. There were some improvements in correlations between the winter and summer rounds of data collection among all risk factors, particularly among speeding observations and in the prevalence of incorrect helmet use (Table 3).

Table 3. Reliability in behavioral risk factor prevalence: Pearson correlation and *P* value.

Risk factor	Winter		Summer	
	Correlation value (*r*)	*P* value	Correlation value (*r*)	*P* value
Helmet use[a]				
Any helmet use	0.98	<.001	0.99	<.001
Correct helmet use	0.82	<.001	0.86	<.001
Incorrect helmet use	0.79	<.001	0.92	<.001
Seatbelt use[a]				
Correct seatbelt use	0.82	<.001	0.82	<.001
Speeding[b]				
No speeding	0.73	<.001	0.80	<.001
≤10 kph[c] over speed limit	0.61	<.001	0.63	<.001
>10 kph over speed limit	0.72	<.001	0.73	<.001

[a]Matched by date, time, location, age, sex, and position.

[b]Matched by date, time, location, vehicle type, and ownership.

[c]kph: kilometers per hour.

Overall Comparison: Risk Factor Proportions

The descriptive analyses showed similar results between methods for most subgroups defined by vehicle and occupant characteristics. For helmet observations, the digital and paper datasets showed correct helmet use proportions within 2 percentage points of each other, across subgroups by age, sex, and role or position. Among seatbelt observations, digital and paper seatbelt use prevalence by vehicle type and occupant position were within 5 percentage points of each other, except for those subgroups with fewer than 200 observations. For comparisons of speeding by vehicle ownership, digital and paper percentages within each category of speeding were also within 5 percentage points of each other. Across all risk factors, the comparability decreased as the subgroup sample size decreased (Multimedia Appendices 2-4).

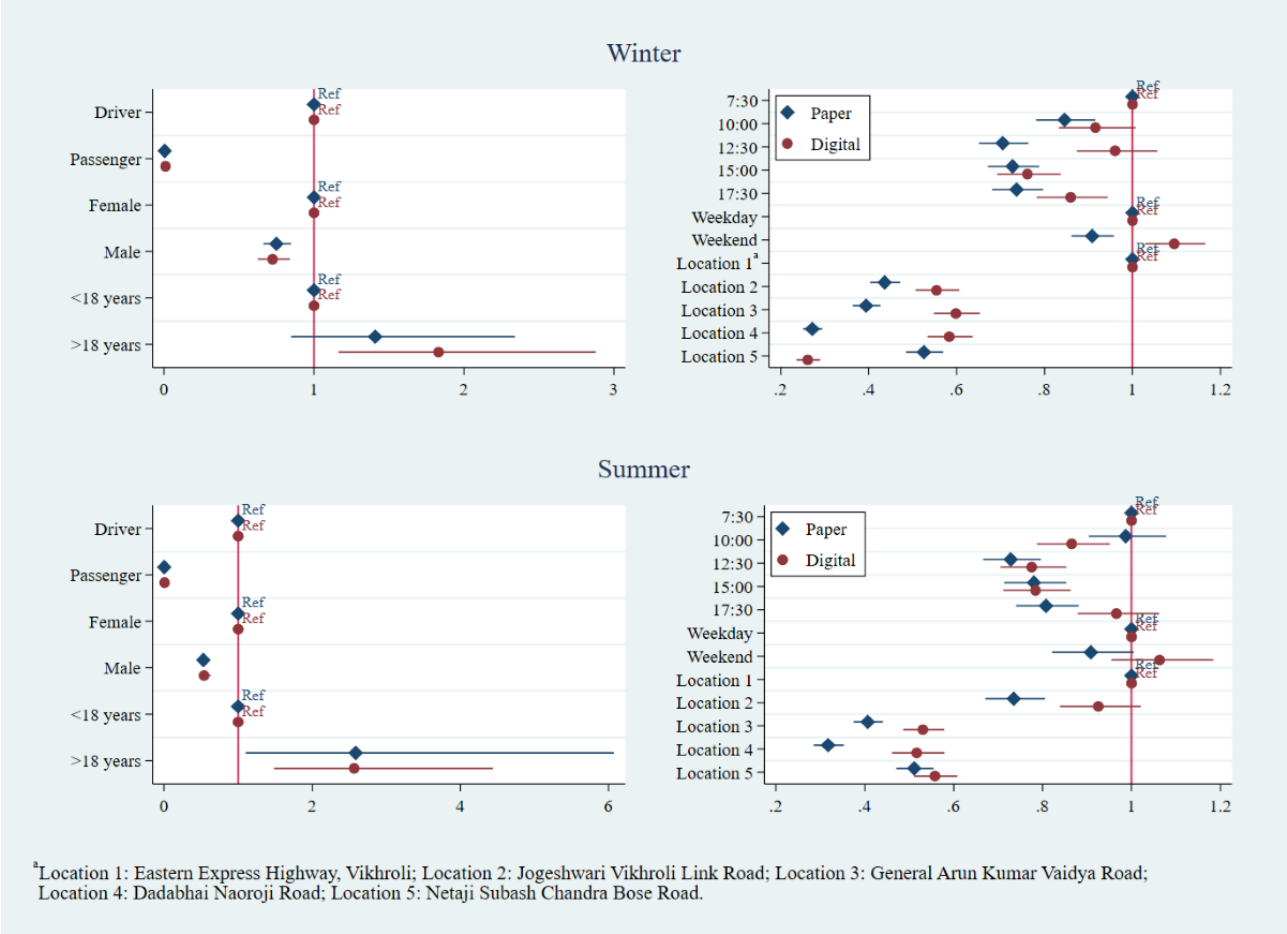
Overall Comparison: Regression Results

The multivariate logistic regressions for correct helmet use, correct seatbelt use, and any speeding showed more variability. Estimates for adjusted odds ratios were largely similar for the associations between occupant characteristics (age, sex, and position) and correct helmet use, regardless of method or round of data collection, adjusting for the day, time, and location (Figure 4). Passengers, compared with drivers, had consistently lower odds of using a helmet correctly, as did males, compared with females, and occupants under 18 years of age, compared with those over 18 years of age. However, the age estimate in the winter 2018 paper dataset was not significant, leading to a different inference from the digital dataset about the impact of age on helmet use between digital and paper. By contrast, the adjusted odds ratios for day of week, time, and location of data collection did not have as many overlapping CIs in either round of helmet use data collection.

As with helmet use data collection, the relationships between occupant characteristics and seatbelt use were comparable between digital and paper datasets, while adjusted odds ratios for day of week, time, and location showed more variability (Figure 5). We regressed sex, age, occupant position, day, time, location, vehicle type, and vehicle ownership onto correct seatbelt use. Seatbelt use was higher among males as compared with females, with front and rear passengers showing very low odds of seatbelt use, as compared with drivers, in both paper and digital datasets. There were similar estimates with overlapping CIs for all occupant age groups, although inferences differed between data collection methods for some age groups (12-17 years and over 60 years) in the winter round of data collection, as the paper data collection CI overlapped with the null value.

Similar to the other two risk factors, in the speeding regressions, the adjusted odds ratios for time showed some variability, controlling for vehicle type, vehicle ownership, and environmental speed deterrents (Figure 6). The effect of speed deterrents, which were only present during the winter round of data collection, showed a similar impact on reducing speeding regardless of data collection method; both the magnitude of the adjusted odds ratios and inferences were similar. Please see Multimedia Appendices 5-7 for further details on regression results for all risk factors.

Figure 4. Correct helmet use: adjusted odds ratios by data collection method and round.



[a]Location 1: Eastern Express Highway, Vikhroli; Location 2: Jogeshwari Vikhroli Link Road; Location 3: General Arun Kumar Vaidya Road; Location 4: Dadabhai Naoroji Road; Location 5: Netaji Subash Chandra Bose Road.

Figure 5. Seatbelt use: adjusted odds ratios by data collection method and round.

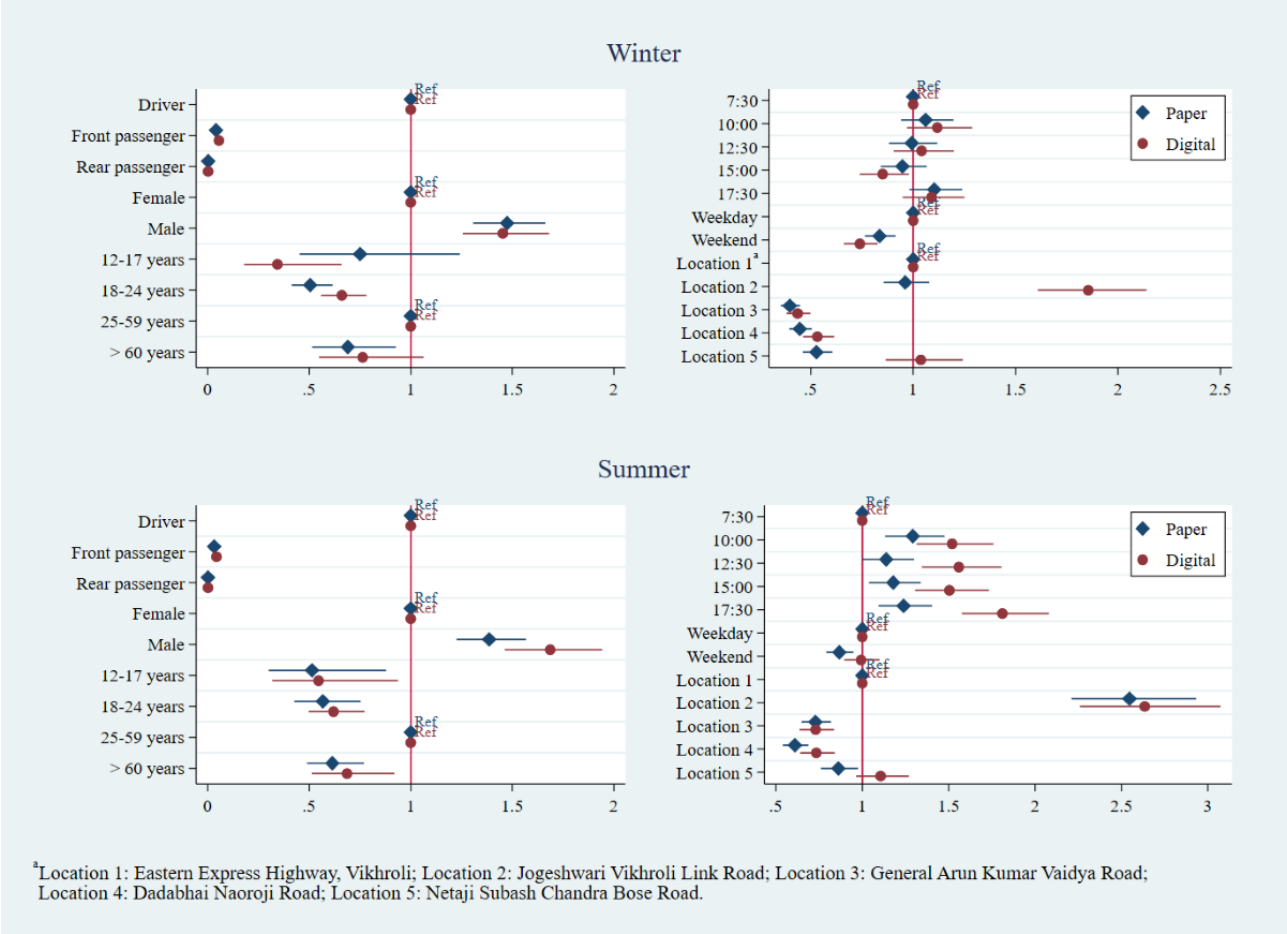
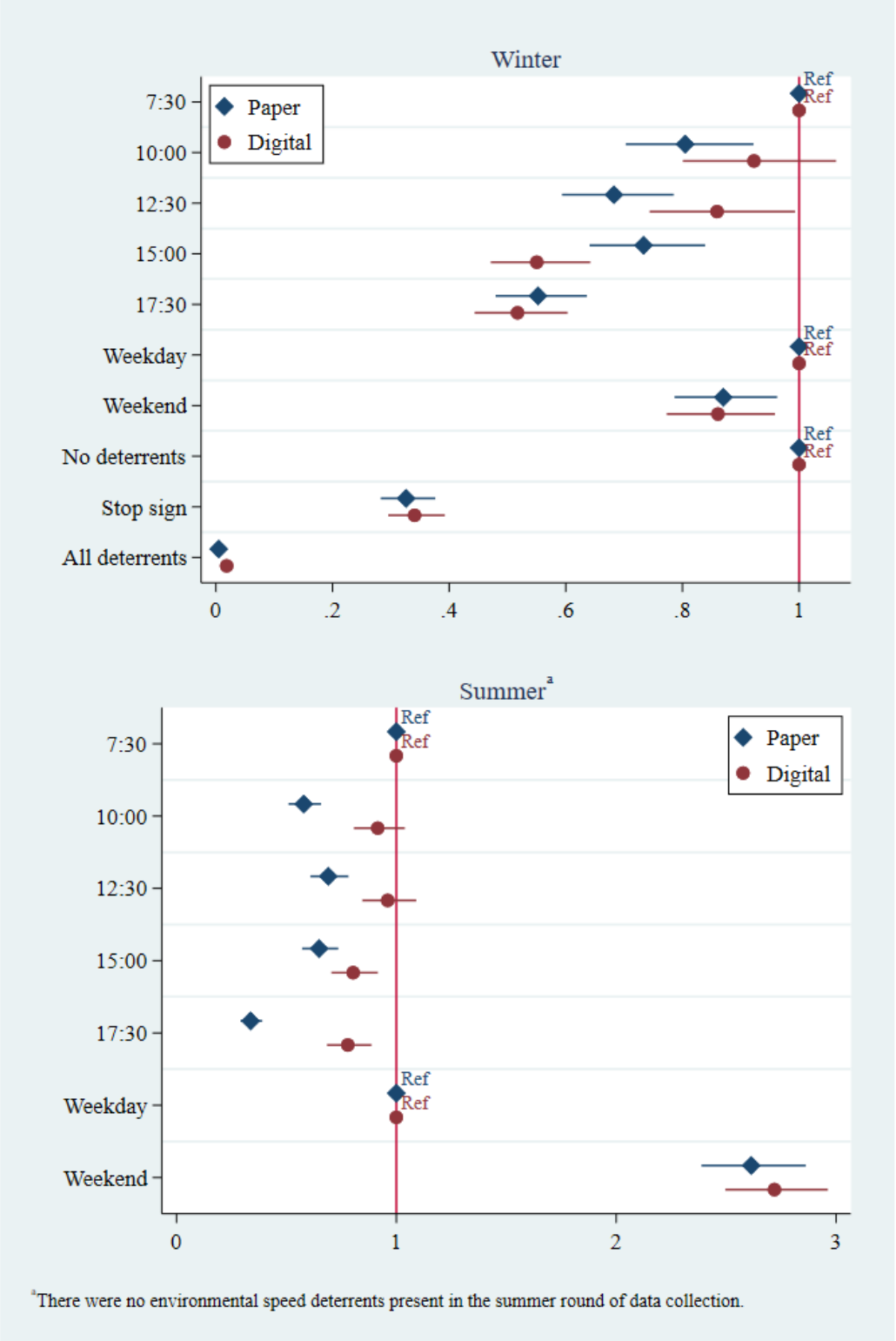


Figure 6. Speeding: adjusted odds ratios by data collection method and round.



Efficiency: Cost Comparison

A comparison of the initial set up and running costs for paper and digital data collection for this study is presented in Table 4. Field data collection costs for staffing and transportation were identical between methods. The field data collection for a single round was covered in 2880 labor hours for each method. Fixed costs or one-time costs, for digital and paper data collection, differed. Four android tablets with sim cards, cases, and power banks were used for digital data collection. For the paper data collection, two desktop computers with hard drives were purchased in the first round, and a laboratory space for paper data entry and storage was rented.

Variable costs or running costs for materials and supplies in paper data collection included paper printing, pens, and clipboards. For digital data collection, variable costs included monthly cellular data plans to ensure continuous connectivity and annual subscription of server space and backup storage.

The variable costs for labor were also different between digital and paper data collection methods. Data management for paper data included data entry, cleaning, and verification. Data entry required 450 labor hours divided over 5 persons, with data cleaning and verification taking an additional 180 labor hours. This could take an additional 26.25 days of work per person, an estimated 105 labor days for 4 data entry operators working 6 hours per day. Digital data management included initial app development, data cleaning, and daily server monitoring, taking approximately 12 labor hours per round.

During the first round, digital data collection cost 1.54% (US $310/20,071) more than paper data collection, driven by the fixed costs of software development and server setup. However, over both rounds in year 1, digital data collection cost 7.73% (US $3011/38,947) less than paper data collection, as with further data collection the higher variable costs in paper exceed the higher fixed costs with digital data collection. Given that the project would potentially last for multiple years, projected costs for data collection over 4 years is presented in Table 5. Over 4 years with two rounds of data collection per year, digital data collection would result in 10.03% (US $15,268/152,203) cost savings, primarily by cutting data entry and verification costs.

Table 4. Cost comparison between paper and digital data collection across rounds of data collection in US dollars.

Cost description	Paper			Digital		
	Units	Winter, US $	Summer, US $	Units	Winter, US $	Summer, US $
Fixed costs						
Hardware						
Computers for data entry, number	2	1195	N/A[a]	N/A	N/A	N/A
Tablets and accessories, number	N/A	N/A	N/A	4	1972	N/A
Others						
Laboratory space for data entry and storage, total cost	1	299	299	N/A	N/A	N/A
App development, labor hours	N/A	N/A	N/A	20	955	N/A
Server and back up, annual total cost	N/A	N/A	N/A	1	1899	N/A
Training						
Total training costs, days	2	299	299	2	299	299
Variable cost						
Data collection personnel						
Observation, labor hours	1260	5647	5647	1260	5647	5647
Recording, labor hours	1260	5647	5647	1260	5647	5647
Supervision, labor hours	360	2690	2690	180	1345	1345
Data Management						
Data entry, labor hours	450	1008	1008	N/A	N/A	N/A
Data cleaning and verification, labor hours	180	1076	1076	12	364	364
Others						
Supplies (pens, paper, etc.), total cost	1	195	195	N/A	N/A	N/A
Tablet data plans, days	N/A	N/A	N/A	360	239	239
Transportation, trips	180	2017	2017	180	2017	2017

[a]N/A: not applicable. This type of cost was not incurred.

Table 5. Realized and projected cost savings by switching to digital data collection in US dollars.

Cost description over years	Winter		Summer		Yearly savings from DDC[a]
	Paper	Digital	Paper	Digital	
Year 1 (realized)					
Fixed costs, US $	1793	5125[b,c]	598	299	N/A[d]
Variable costs, US $	18,278	15,256	18,278	15,256	N/A
Total costs, US $	20,071	20,381	18,876	15,555	N/A
Cost savings from DDC[e] (US $), n (%)	N/A	–310 (–1.54)	N/A	3321 (17.59)	3011 (7.73)
Year 2 (projected)					
Fixed costs, US $	598	2198[b]	598	299	N/A
Variable costs, US $	18,278	15,256	18,278	15,256	N/A
Total costs, US $	18,876	17,454	18,876	15,555	N/A
Cost savings from DDC[e] (US $), n (%)	N/A	1422 (7.53)	N/A	3321 (17.59)	4743 (12.56)
Year 3 (projected)					
Fixed costs, US $	598	4170[b,f]	598	299	N/A
Variable costs, US $	18,278	15,256	18,278	15,256	N/A
Total costs, US $	18,876	19,426	18,876	15,555	N/A
Cost savings from DDC[e] (US $), n (%)	N/A	–550 (–2.91)	N/A	3321 (17.59)	2771 (7.34)
Year 4 (projected)					
Fixed costs, US $	598	2198[b]	598	299	N/A
Variable costs, US $	18,278	15,256	18,278	15,256	N/A
Total costs, US $	18,876	17,454	18,876	15,555	N/A
Cost savings from DDC[e] (US $), n (%)	N/A	1422 (7.53)	N/A	3321 (17.59)	4743 (12.56)
Total cost savings (projected; US $), n (%)	N/A	N/A	N/A	N/A	15,268 (10.03)

[a]DDC: digital data collection.

[b]Includes yearly subscription of the cloud server.

[c]Includes initial app development.

[d]N/A: not applicable.

[e]This is difference between paper and digital costs as a percentage of the paper cost ([paper–digital]/paper).

[f]Assumes replacement of tablets and accessories after 2 years.

Efficiency: Time Comparison

As mentioned earlier, productivity was higher in paper collection as compared with digital data collection, with improvement in the correlation of volumes over time. With observation periods lasting 90 min in each data collection method, this higher productivity in paper data collection is due to the fact that it took field teams between 2-6 seconds longer to collect data digitally than to collect data using paper forms, on average (Table 6). Such small differences added up over the several hundred observations made per observation period and over 75 observation periods per round. Interestingly, the improvement in the correlation of productivity between the winter and summer rounds was due not only to an increase in speed in digital data collection but a slight drop in speed in paper data collection. Between the winter and summer data collection rounds, the time to complete an observation decreased or remained the same in digital but increased slightly in paper.

Although each observation took less time in paper data collection, it took less time to obtain a complete dataset with digital data collection. The time for training and field data collection were the same between digital and paper data collection methods, with 2 days for training and 15 days for field data collection. Data entry and verification in paper data took an additional 630 hours, or 26.25 days for 4 data entry operators to complete, which was done concurrently with ongoing fieldwork. Under the best conditions, this would delay receipt of data collection by a minimum of 11 days after fieldwork was complete. By contrast, digital data collection initially required 20 hours before fieldwork to develop the app and set up the server, and an average of 12 hours of data cleaning following receipt of the data before it was ready for

analysis. Switching from paper to digital data collection reduced the time to receive a clean dataset after fieldwork was completed from 630 hours to 12 hours, or from 11 days to 2 days.

Table 6. Time in seconds to complete an observation by data collection method and round.

Risk factor	Winter, time (seconds)		Summer, time (seconds)	
	Paper	Digital	Paper	Digital
Helmet use[a]	12.0	17.6	14.6	17.4
Seatbelt use[a]	11.3	17.5	12.4	16.4
Speeding[b]	13.4	16.7	14.3	16.7

[a]Observation on a vehicle occupant.

[b]Observation on a vehicle.

Discussion

Productivity

In our study, productivity was higher in paper data collection as compared with digital data collection regardless of the round of data collection or risk factor. The majority of studies comparing productivity between digital and paper data collection methods found equivalent or less time to complete a digital survey or observation, as compared with paper data collection [12-16]. Paper data collection may have taken less time in our study for several reasons. First, the data collectors had greater familiarity and experience with paper data collection, having previously completed five rounds of paper-based data collection over 3 years. Second, the act of swiping through multiple screens during digital data collection may take more time than filling in a single row in a predesigned table format on paper. Third, the digital tools used logic checks and constraints to prevent errors, which may delay recording the response. Although digital productivity did improve over time, it is not possible to project whether it would ever eventually catch up to the productivity seen in paper data collection, due to the formatting of the digital module and the logic checks imposed. Researchers and project managers may wish to spend additional time on classroom and field training with digital data collection tools, to give field staff the extra practice required to close this productivity gap as much as possible.

Despite the higher productivity in paper data collection, both digital and paper data collection methods were able to obtain a large enough sample size to estimate a proportion within ± 1 percentage point, which we judged to be a reasonable margin of error. Moreover, this higher productivity in paper did not translate into higher quality statistical models, measured using AIC.

Reliability

Similar to other direct comparisons of digital and paper data collection, we found high reliability between the two methods [15,17-27]. This high reliability was seen across risk factors, and improved between rounds of data collection, perhaps indicating improvements in data collector facility with the digital modules over time.

Overall Comparison: Risk Factor Proportions and Regression Results

The descriptive analyses showed comparable results of behavioral risk factor prevalence between subgroups defined by occupant, vehicle, or environmental characteristics, but with comparability decreasing with sample size. Other studies that conducted descriptive analyses also found similar distributions between data collection methods [12,15-28].

However, the multivariable regression analyses showed that the digital and paper datasets produced comparable results for some variables but not for others. Occupant characteristics—age, sex, and position—showed similar adjusted odds ratios, meaning that a researcher using digital data collection would have come to the same conclusions as a researcher using paper-based data collection when trying to understand the occupant-related factors associated with the use of protective equipment. However, there was less comparability between digital and paper methods in odds ratios for vehicle-related factors, or for day, time, and location, across risk factors.

Owing to smaller sample sizes in some subgroups, some age categories, vehicle types, and vehicle ownership types showed different prevalence estimates between paper and digital methods. Differences in regression results by day, time, and location, between digital and paper methods seem to be due to the nature of observational sampling and the differences in productivity between digital and paper methods. With faster data collection in paper, digital and paper data collectors do not make observations on the *same* target vehicles and their occupants. Such random differences are less pronounced for occupant characteristics (age, sex, or position), or when the prevalence of the risk factor are significantly modified by sex and occupant roles (drivers vs passengers). Differences in risk factor prevalence are less pronounced among the different locations and times of day, and therefore, random differences in target vehicle selection may have a larger effect on prevalence and odds ratio estimates. These results cannot be explained by surveyor effects, as the pool of data collectors were randomly assigned to data collection teams, and teams were randomly switched between methods and among locations, times of day, and days of the week.

Efficiency: Cost Comparison

Digital data collection had higher fixed costs and lower variable costs compared with paper data collection. In the initial round

of data collection, digital data collection was slightly more expensive, driven by the costs of purchasing tablets, but became less expensive with additional rounds of data collection. Digital data collection is significantly more cost efficient than paper when costs are annualized over several years with ongoing data collection. Digital data collection will have the greatest cost efficiency when used by projects with larger sample sizes, or multiple and extended rounds of data collections [29]. As labor, training, and transportation cost would be the same between methods, regardless of annual salary increase or inflation rate, the cost savings by digital data collection are driven primarily by savings in data management, particularly data entry. Other studies have similarly found higher fixed costs for digital data collection and lower variable costs, providing increased cost savings for larger sample sizes [13,14,16,30-33].

Cost differences between digital and paper will vary by project and context. The relative cost of labor, hardware, and electricity may differ by geographic location. Additionally, the cost of data collection platforms differs by company and by project. For example, for this project, the server space was purchased to accommodate data for all 10 cities, and hence the actual space required for Mumbai data was smaller and could have cost significantly less. Contractual agreements and price negotiations for server space could be influenced by a number of factors such as the size of the company, infrastructure and service details, promotions, packages, size and configuration of the server space, program, data security and back up arrangements, length of time, region, and location. The cost figures provided are particular to this project, and we encourage researchers for whom cost differential is a major consideration to draw up similar comparative budgets for digital and paper data collection methods for a given project. On the basis of our experience and the existing literature, for each study, there will be a sample size threshold above which digital data collection will provide cost savings, and below which paper data collection will be the more cost-effective option.

Our study was not able to directly measure a third dimension of efficiency: accuracy, or the reduction in errors achieved per resource unit expended. However, we note that the error protection built into the digital app through logic checks and constraints had two advantages. It decreased the cost of digital data collection as compared with paper, as evident by significantly fewer hours spent on verification and cleaning of data. It also decreased the size of the AIC seen in digital data collection compared with paper, even with lower sample size in digital methods. In other words, digital data collection provided our project with more precision at a lower cost.

Efficiency: Time Comparison

One of the biggest benefits of digital data collection is the real-time nature of data collection and availability: once data collection starts, the digital data are immediately available for data cleaning and analysis, while researchers using paper data collection must wait for data entry and reconciliation to be complete. It is important to note that for complex questionnaires, more time may be needed at startup for digital data collection as compared with paper-based data collection [24,34]. However, the efficiency for subsequent rounds of data collection would be significantly increased. The majority of studies comparing digital and paper data collection methods found that digital data collection reduced the time to obtain a complete dataset, as they avoided double data entry and reconciliation [14,16,33,35].

Conclusions

Our study did not face trade-offs among time efficiency, cost efficiency, statistical reliability, and descriptive comparability when deciding between digital and paper methods, as digital data collection proved equivalent or superior on these domains in the context of our project. Digital data collection provides estimates which are precise and reliable with paper data collection, with overall comparable results. Digital data collection may take longer per survey but takes less calendar time to obtain a completed dataset. The setting, context, length of the survey, and desired sample size play a role in determining the extent of time and cost efficiency. However, we may have encountered a method effect due to sampling methods for observational studies and random error.

Researchers considering using digital data collection may benefit from developing comparative budgets for each of the two methods, and pilot testing each method to understand relative productivity. When working with data collectors who are unfamiliar with handheld devices or with the chosen digital data collection app, additional classroom and field training on the digital data collection tools would improve productivity. Although this may slightly narrow the time efficiency advantage that digital data collection has over paper and increase the cost of training for digital methods as compared with paper, this additional effort would be worthwhile especially in studies relying on higher productivity. As trade-offs among cost, timeliness, and comparability and the relative importance of each could be unique to every data collection project, researchers should carefully consider the questionnaire complexity, target sample size, implementation plan, cost and logistical constraints, and geographical contexts when making the decision between digital and paper [36].

Authors' Contributions

NT and AM conceptualized the study and prepared all manuscript drafts. AM developed the digital data collection tool. NT led the data analysis and prepared results. PV and SG led the data collection in Mumbai. RP and AM participated in training and implementation of digital data collection in Mumbai. AB provided oversight of the study and critical feedback in data analysis and manuscript drafts. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Akaike Information Criterion for logistic regressions, by risk factor, round, and method of data collection.
[DOCX File , 14 KB - jmir_v22i5e17129_app1.docx]

Multimedia Appendix 2
Whole sample proportions of correct, incorrect and no helmet use by occupant role, age, and sex.
[PNG File , 102 KB - jmir_v22i5e17129_app2.png]

Multimedia Appendix 3
Whole sample proportions of restraint use by occupant role and vehicle type.
[PNG File , 107 KB - jmir_v22i5e17129_app3.png]

Multimedia Appendix 4
Whole sample proportion of vehicles speeding by vehicle ownership.
[PNG File , 96 KB - jmir_v22i5e17129_app4.png]

Multimedia Appendix 5
Correct helmet use: adjusted odds ratios and 95% CIs by round and method of data collection.
[DOCX File , 17 KB - jmir_v22i5e17129_app5.docx]

Multimedia Appendix 6
Seatbelt use: adjusted odds ratios and 95% CIs by round and method of data collection.
[DOCX File , 20 KB - jmir_v22i5e17129_app6.docx]

Multimedia Appendix 7
Speeding: adjusted odds ratios and 95% CIs by round and method of data collection.
[DOCX File , 18 KB - jmir_v22i5e17129_app7.docx]

References

1. World Health Organization. Geneva: World Health Organization; 2018. Global Status Report on Road Safety, 2018 URL: https://www.who.int/publications-detail/global-status-report-on-road-safety-2018 [accessed 2019-07-29]
2. World Health Organization. 2011. Global Plan for the Decade of Action for Road Safety 2011-2020 URL: https://www.who.int/roadsafety/decade_of_action/plan/plan_english.pdf?ua=1 [accessed 2019-07-29]
3. Bloomberg Philanthropies. 2018. Bloomberg Philanthropies Initiative for Global Road Safety URL: https://www.bloomberg.org/program/public-health/road-safety/ [accessed 2018-12-03]
4. Mehmood A, Taber N, Bachani AM, Gupta S, Paichadze N, Hyder AA. Paper Versus Digital Data Collection for Road Safety Risk Factors: Reliability Comparative Analysis From Three Cities in Low- and Middle-Income Countries. J Med Internet Res 2019 May 28;21(5):e13222-e13264 [FREE Full text] [doi: 10.2196/13222] [Medline: 31140431]
5. Raghavan C. Encyclopedia Britannica. Mumbai URL: https://www.britannica.com/place/Mumbai [accessed 2019-09-13]
6. World Population Review. 2019. Mumbai Population 2020 URL: http://worldpopulationreview.com/world-cities/mumbai-population/ [accessed 2019-09-13]
7. Government of India. New Delhi; 2016. Service Area-Wise Internet Subscriber-Base and Internet Subscribers Per 100 Population as on 30th June 2016 (From: Ministry of Communications) URL: https://data.gov.in/resources/service-area-wise-internet-subscriber-base-and-internet-subscribers-100-population-30th [accessed 2019-09-13]
8. Census 2011. 2011. Mumbai (Greater Mumbai) City Census 2011 Data URL: https://www.census2011.co.in/census/city/365-mumbai.html [accessed 2020-03-17]
9. Ministry of Road Transport & Highways. 2017. Road Accidents in India - 2017 URL: http://morth.nic.in/sites/default/files/Road_Accidents_in_India_2017.pdf [accessed 2019-08-21]
10. Ministry of Statistics and Program Implementation. New Delhi: Ministry of Statistics and Program Implementation, Government Of India; 2017. Total Registered Motor Vehicles in Million Plus Cities URL: http://mospi.nic.in/statistical-year-book-india/2017/189 [accessed 2019-08-21]
11. KoBoToolbox.: Harvard Humanitarian Initiative URL: https://www.kobotoolbox.org/ [accessed 2018-12-03]
12. Vivoda JM, Eby DW. Using personal digital assistants (PDAs) for the collection of safety belt use data in the field. Behav Res Methods 2006 Feb;38(1):158-164. [doi: 10.3758/brm.38.1.158] [Medline: 16817526]
13. Zhang S, Wu Q, van Velthoven MH, Chen L, Car J, Rudan I, et al. Smartphone versus pen-and-paper data collection of infant feeding practices in rural China. J Med Internet Res 2012 Sep 18;14(5):e119 [FREE Full text] [doi: 10.2196/jmir.2183] [Medline: 22989894]

14. Dillon DG, Pirie F, Rice S, Pomilla C, Sandhu MS, Motala AA, African Partnership for Chronic Disease Research (APCDR). Open-source electronic data capture system offered increased accuracy and cost-effectiveness compared with paper methods in Africa. J Clin Epidemiol 2014 Dec;67(12):1358-1363 [FREE Full text] [doi: 10.1016/j.jclinepi.2014.06.012] [Medline: 25135245]
15. Pawar SG, Ramani PS, Prasad A, Dhar A, Babhulkar SS, Bahurupi YA. Software version of Roland Morris Disability questionnaire for outcome assessment in low back pain. Neurol Res 2017 Apr;39(4):292-297. [doi: 10.1080/01616412.2017.1297555] [Medline: 28337948]
16. King JD, Buolamwini J, Cromwell EA, Panfel A, Teferi T, Zerihun M, et al. A novel electronic data collection system for large-scale surveys of neglected tropical diseases. PLoS One 2013;8(9):e74570 [FREE Full text] [doi: 10.1371/journal.pone.0074570] [Medline: 24066147]
17. Bernabe-Ortiz A, Curioso WH, Gonzales MA, Evangelista W, Castagnetto JM, Carcamo CP, et al. Handheld computers for self-administered sensitive data collection: a comparative study in Peru. BMC Med Inform Decis Mak 2008 Mar 19;8:11 [FREE Full text] [doi: 10.1186/1472-6947-8-11] [Medline: 18366687]
18. McBride JS, Anderson RT, Bahnson JL. Using a hand-held computer to collect data in an orthopedic outpatient clinic: a randomized trial of two survey methods. Med Care 1999 Jul;37(7):647-651. [doi: 10.1097/00005650-199907000-00004] [Medline: 10424635]
19. Jaspan HB, Flisher AJ, Myer L, Mathews C, Seebregts C, Berwick JR, et al. Brief report: methods for collecting sexual behaviour information from South African adolescents--a comparison of paper versus personal digital assistant questionnaires. J Adolesc 2007 Apr;30(2):353-359. [doi: 10.1016/j.adolescence.2006.11.002] [Medline: 17187853]
20. Fletcher LA, Erickson DJ, Toomey TL, Wagenaar AC. Handheld computers: A feasible alternative to paper forms for field data collection. Eval Rev 2003 Apr;27(2):165-178. [doi: 10.1177/0193841X02250527] [Medline: 12703341]
21. Saleh KJ, Radosevich DM, Kassim RA, Moussa M, Dykes D, Bottolfson H, et al. Comparison of commonly used orthopaedic outcome measures using palm-top computers and paper surveys. J Orthop Res 2002 Nov;20(6):1146-1151 [FREE Full text] [doi: 10.1016/S0736-0266(02)00059-1] [Medline: 12472221]
22. Junker U, Freynhagen R, Längler K, Gockel U, Schmidt U, Tölle TR, et al. Paper versus electronic rating scales for pain assessment: a prospective, randomised, cross-over validation study with 200 chronic pain patients. Curr Med Res Opin 2008 Jun;24(6):1797-1806. [doi: 10.1185/03007990802121059] [Medline: 18485269]
23. Byrom B, Doll H, Muehlhausen W, Flood E, Cassedy C, McDowell B, et al. Measurement equivalence of patient-reported outcome measure response scale types collected using bring your own device compared to paper and a provisioned device: results of a randomized equivalence trial. Value Health 2018 May;21(5):581-589 [FREE Full text] [doi: 10.1016/j.jval.2017.10.008] [Medline: 29753356]
24. Ahmed R, Robinson R, Elsony A, Thomson R, Squire SB, Malmborg R, et al. A comparison of smartphone and paper data-collection tools in the Burden of Obstructive Lung Disease (BOLD) study in Gezira state, Sudan. PLoS One 2018;13(3):e0193917 [FREE Full text] [doi: 10.1371/journal.pone.0193917] [Medline: 29518132]
25. Goldstein LA, Gibbons MB, Thompson SM, Scott K, Heintz L, Green P, et al. Outcome assessment via handheld computer in community mental health: consumer satisfaction and reliability. J Behav Health Serv Res 2011 Jul;38(3):414-423 [FREE Full text] [doi: 10.1007/s11414-010-9229-4] [Medline: 21107916]
26. Richter JG, Becker A, Koch T, Nixdorf M, Willers R, Monser R, et al. Self-assessments of patients via Tablet PC in routine patient care: comparison with standardised paper questionnaires. Ann Rheum Dis 2008 Dec;67(12):1739-1741. [doi: 10.1136/ard.2008.090209] [Medline: 18647853]
27. VanDenKerkhof EG, Goldstein DH, Blaine WC, Rimmer MJ. A comparison of paper with electronic patient-completed questionnaires in a preoperative clinic. Anesth Analg 2005 Oct;101(4):1075-80, table of contents. [doi: 10.1213/01.ane.0000168449.32159.7b] [Medline: 16192524]
28. Drummond HE, Ghosh S, Ferguson A, Brackenridge D, Tiplady B. Electronic quality of life questionnaires: a comparison of pen-based electronic questionnaires with conventional paper in a gastrointestinal study. Qual Life Res 1995 Feb;4(1):21-26. [doi: 10.1007/bf00434379] [Medline: 7711686]
29. Thakkar M, Floretta J, Dhar D, Wilmink N, Sen S, Keleher N, et al. The Clear Initiative. 2016. Mobile-Based Technology for Monitoring & Evaluation URL: https://www.theclearinitiative.org/sites/clearinitiative/files/2016-04/mobile-based-tech.pdf [accessed 2019-08-21]
30. Seebregts CJ, Zwarenstein M, Mathews C, Fairall L, Flisher AJ, Seebregts C, et al. Handheld computers for survey and trial data collection in resource-poor settings: development and evaluation of PDACT, a Palm Pilot interviewing system. Int J Med Inform 2009 Nov;78(11):721-731. [doi: 10.1016/j.ijmedinf.2008.10.006] [Medline: 19157967]
31. Mukasa O, Mushi HP, Maire N, Ross A, de Savigny D. Do surveys with paper and electronic devices differ in quality and cost? Experience from the Rufiji Health and demographic surveillance system in Tanzania. Glob Health Action 2017;10(1):1387984 [FREE Full text] [doi: 10.1080/16549716.2017.1387984] [Medline: 29157182]
32. Giduthuri JG, Maire N, Joseph S, Kudale A, Schaetti C, Sundaram N, et al. Developing and validating a tablet version of an illness explanatory model interview for a public health survey in Pune, India. PLoS One 2014;9(9):e107374 [FREE Full text] [doi: 10.1371/journal.pone.0107374] [Medline: 25233212]

33. Yu P, de Courten M, Pan E, Galea G, Pryor J. The development and evaluation of a PDA-based method for public health surveillance data collection in developing countries. Int J Med Inform 2009 Aug;78(8):532-542. [doi: 10.1016/j.ijmedinf.2009.03.002] [Medline: 19369114]
34. Byass P, Hounton S, Ouédraogo M, Somé H, Diallo I, Fottrell E, et al. Direct data capture using hand-held computers in rural Burkina Faso: experiences, benefits and lessons learnt. Trop Med Int Health 2008 Jul;13(Suppl 1):25-30 [FREE Full text] [doi: 10.1111/j.1365-3156.2008.02084.x] [Medline: 18578809]
35. Inman-Narahari F, Giardina C, Ostertag R, Cordell S, Sack L. Digital data collection in forest dynamics plots. Methods Ecol Evol 2010 Sep;1(3):274-279. [doi: 10.1111/j.2041-210x.2010.00034.x]
36. Tomkys E, Lombardini S. Going Digital: Using Digital Technology to Conduct Oxfam's Effectiveness Reviews. Oxford: Oxfam GB; Sep 2015.

Abbreviations

AIC: Akaike Information Criteria
BIGRS: Bloomberg Initiative for Global Road Safety
kph: kilometers per hour
LMICs: low- and middle-income countries
RTI: road traffic injury

Edited by G Eysenbach; submitted 08.12.19; peer-reviewed by M Peden, P Banik; comments to author 15.02.20; revised version received 17.02.20; accepted 26.02.20; published 22.05.20.

Please cite as:
Taber N, Mehmood A, Vedagiri P, Gupta S, Pinto R, Bachani AM
Paper Versus Digital Data Collection Methods for Road Safety Observations: Comparative Efficiency Analysis of Cost, Timeliness, Reliability, and Results
J Med Internet Res 2020;22(5):e17129
URL: http://www.jmir.org/2020/5/e17129/
doi:10.2196/17129
PMID:32348273

Original Paper

Characterizing the US Population by Patterns of Mobile Health Use for Health and Behavioral Tracking: Analysis of the National Cancer Institute's Health Information National Trends Survey Data

Camella J Rising[1], PhD, MS, RDN; Roxanne E Jensen[2], PhD; Richard P Moser[3], PhD; April Oh[1], PhD, MPH

[1]Health Communication and Informatics Research Branch, Behavioral Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Rockville, MD, United States

[2]Outcomes Research Branch, Healthcare Delivery Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Rockville, MD, United States

[3]Office of the Associate Director, Behavioral Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Rockville, MD, United States

Corresponding Author:
Camella J Rising, PhD, MS, RDN
Health Communication and Informatics Research Branch, Behavioral Research Program
Division of Cancer Control and Population Sciences
National Cancer Institute
9606 Medical Center Drive, 3E636
Rockville, MD, 20850
United States
Phone: 1 240 276 5262
Email: camella.rising@nih.gov

Abstract

Background: Multiple types of mobile health (mHealth) technologies are available, such as smartphone health apps, fitness trackers, and digital medical devices. However, despite their availability, some individuals do not own, do not realize they own, or own but do not use these technologies. Others may use mHealth devices, but their use varies in tracking health, behaviors, and goals. Examining patterns of mHealth use at the population level can advance our understanding of technology use for health and behavioral tracking. Moreover, investigating sociodemographic and health-related correlates of these patterns can provide direction to researchers about how to target mHealth interventions for diverse audiences.

Objective: The aim of this study was to identify patterns of mHealth use for health and behavioral tracking in the US adult population and to characterize the population according to those patterns.

Methods: We combined data from the 2017 and 2018 National Cancer Institute Health Information National Trends Survey (N=6789) to characterize respondents according to 5 mutually exclusive reported patterns of mHealth use for health and behavioral tracking: (1) mHealth nonowners and nonusers report not owning or using devices to track health, behaviors, or goals; (2) supertrackers track health or behaviors and goals using a smartphone or tablet plus other devices (eg, Fitbit); (3) app trackers use only a smartphone or tablet; (4) device trackers use only nonsmartphone or nontablet devices and do not track goals; and (5) nontrackers report having smartphone or tablet health apps but do not track health, behaviors, or goals.

Results: Being in the mHealth nonowners and nonusers category (vs all mHealth owners and users) is associated with males, older age, lower income, and not being a health information seeker. Among mHealth owners and users, characteristics of device trackers and supertrackers were most distinctive. Compared with supertrackers, device trackers have higher odds of being male (odds ratio [OR] 2.22, 95% CI 1.55-3.19), older age (vs 18-34 years; 50-64 years: OR 2.83, 95% CI 1.52-5.30; 65+ years: OR 6.28, 95% CI 3.35-11.79), have an annual household income of US $20,000 to US $49,999 (vs US $75,000+: OR 2.31, 95% CI 1.36-3.91), and have a chronic condition (OR 1.69, 95% CI 1.14-2.49). Device trackers also have higher odds of not being health information seekers than supertrackers (OR 2.98, 95% CI 1.66-5.33).

Conclusions: Findings revealed distinctive sociodemographic and health-related characteristics of the population by pattern of mHealth use, with notable contrasts between those who do and do not use devices to track goals. Several characteristics of individuals who track health or behaviors but not goals (device trackers) are similar to those of mHealth nonowners and nonusers.

Our results suggest patterns of mHealth use may inform how to target mHealth interventions to enhance reach and facilitate healthy behaviors.

(J Med Internet Res 2020;22(5):e16299) doi:10.2196/16299

KEYWORDS

mobile health; population health; health communication; survey methodology; mobile applications; devices

Introduction

Background

Mobile health (mHealth) includes the use of portable digital devices, such as smartphones, tablet computers, and fitness and medical wearables, to support health. Approximately 80% of US adults own a smartphone [1], which typically has one or more health-related software apps (*health apps*) preinstalled at the point of purchase. Close to 60% of smartphone owners also report installing (ie, downloading) one or more health apps onto their smartphone [2]. Wearable devices with a multitude of sensor types designed to monitor health indicators, such as activity, sleep, blood glucose, and blood pressure, have also penetrated the US marketplace [3,4]. Within the general population, wearable activity monitors (eg, Fitbit) are among the most commonly owned wearable devices and are reportedly used by about 13% of Americans [5].

Despite their ubiquity, ownership and use of diverse types of mHealth technologies for health and behavioral tracking varies. Reported use of mHealth technologies includes communicating with health care providers [6], making dietary or physical activity decisions [7,8], achieving health goals such as weight loss [9], and monitoring chronic conditions such as diabetes or hypertension [10,11]. Some mHealth intervention studies have reported that the use of mHealth technologies is associated with improved health or behavioral outcomes. These include, for example, improvements in diet and physical activity [12,13], weight loss or maintenance and blood glucose reduction [11], and higher quality patient-provider communication [14]. Although these studies suggest the potential for mHealth to enhance patient-provider communication, improve health outcomes, and facilitate behavior change, research examining how mHealth technologies are used by the US population for health and behavioral tracking is needed to understand how to target and tailor mHealth interventions to enhance their reach and efficacy across population subgroups.

Currently, there are gaps in the literature related to understanding use of mHealth technologies for health and behavioral tracking that need to be filled to advance mHealth intervention science. First, most intervention studies to date evaluate only one type of mHealth technology to address a desired health outcome or target behavior [8,15-18]. Yet, as mHealth technologies diversify in functionality, individuals have also become increasingly diverse in the types of mHealth technologies they own and use, with multiuse becoming more common [19]. Therefore, identifying the types of mHealth technologies the US population uses for health or behavioral tracking is an essential component of advancing the science.

Second, there is limited research about whether people in the United States set health goals when they use mHealth technologies for health or behavioral tracking. Goal setting can promote higher and sustained engagement with mHealth interventions [20-22]. Moreover, goal setting is strongly associated with favorable health behavior outcomes, such as increased physical activity and healthy eating in overweight and obese adults [23] and is a characteristic of efficacious mHealth interventions to improve diet, physical activity, and sedentary behaviors of children and adults in the general population [13]. Investigating whether tracking includes goal setting and assessing the characteristics of those who set these goals may increase understanding about who may be more (or less) likely to change health behavior as a result of using mHealth technologies for health or behavioral tracking and facilitate more precise targeting and tailoring of future mHealth interventions.

Finally, past studies have found that many people in the United States report not owning or using mHealth technologies for health or behavioral tracking, and that these individuals differ from mHealth technology owners and users on several characteristics such as age, gender, education, and income [5,24]. However, mHealth technologies for health and behavioral tracking continue to evolve and are adopted at different rates across population subgroups. Thus, continuing to describe and characterize mHealth nonowners and nonusers is necessary to understand patterns of mHealth use for health and behavioral tracking in the US population.

Study Aims

This paper has 3 primary aims. The first is to address gaps in the literature by describing patterns of mHealth use for health and behavioral tracking in the US population. We account for 3 factors in our conceptualization of pattern of mHealth use: (1) whether mHealth technologies are owned and used for health or behavioral tracking, (2) the types of mHealth technologies owned and used for health or behavioral tracking, namely smartphones or tablets and other digital devices such as fitness trackers or medical devices, and (3) whether health or behavioral tracking with mHealth technologies includes goal setting.

On the basis of these 3 factors, we distinguish between 5 mutually exclusive categories, or population subgroups, of mHealth owners and users and mHealth nonowners and nonusers in the United States. *mHealth nonowners and nonusers* are those who report that they do not own or use mHealth technologies for health or behavioral tracking. Among mHealth owners and users, *supertrackers* are those who report using multiple devices—a smartphone or tablet and another device, such as a fitness tracker or medical device—to track health or behaviors and goals, whereas *app trackers* report only using a smartphone

or tablet for the same purpose. *Device trackers* are those who report only using a device other than a smartphone or tablet to track health or behaviors and do not use smartphone or tablet health apps to track goals. Finally, *nontrackers* are those who report only having smartphone or tablet health apps but do not use them to track goals and do not track health or behaviors with other devices.

The second aim of this paper is to describe the characteristics of mHealth nonowners and nonusers by comparing them with people who own or use mHealth technologies for health and behavioral tracking. We consider individual-level factors such as sociodemographics—age, gender, race and ethnicity, education, income, and geographical area. We also consider factors associated with health and health behaviors, such as weight status, having a chronic condition (eg, diabetes), perceived health status, health self-efficacy, and being a health information seeker.

The third and final aim of this paper is to describe and compare the characteristics (sociodemographics, health, and health behaviors) of mHealth owners and users by pattern of mHealth use for health and behavioral tracking. Comparing mHealth owners and users by pattern of mHealth use may provide important insights about those who not only own and use different types of mHealth technologies in the United States but also use those technologies in ways that can promote improved health and health behavior outcomes (ie, goal setting). As mHealth technology development and adoption rapidly evolves, characterizing the population by pattern of mHealth use may inform how to target and tailor mHealth interventions to reach diverse US audiences and facilitate healthy behaviors.

Methods

Data

Data were merged from the National Cancer Institute (NCI)'s Health Information National Trends Survey (HINTS), HINTS 5, Cycle 1, and HINTS 5, Cycle 2. HINTS is a probability-based, cross-sectional survey of the US adult, civilian, noninstitutionalized population. Survey data were collected by paper-and-pencil self-administered questionnaires completed from January through May 2017 (Cycle 1; see Multimedia Appendix 1) and January through May 2018 (Cycle 2, see Multimedia Appendix 2). HINTS has a two-stage, stratified sample; addresses were randomly selected from a US Postal Service file of residential addresses, and a random individual respondent was selected from each sample household (HINTS 5, Cycle 1 response rate=32.4%, N=3285; HINTS 5, Cycle 2 response rate=32.9%, N=3504). Details related to HINTS methodology have been described elsewhere [25].

Sociodemographic and Health-Related Characteristics

Items included in the analysis pertaining to respondent characteristics were asked of the full, merged sample from 2017 and 2018 (N=6789). Sociodemographic variables included in analyses were gender (male and female); age (18-34, 35-49, 50-64, and ≥65 years old); race (white, black, and other race, which combines low-frequency responses for American Indian/Alaska Native, Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, other Asian, Native Hawaiian, Guamanian or Chamorro, Samoan, and other Pacific Islander); Hispanic ethnicity (Hispanic, non-Hispanic); education (less than high school or high school graduate, technical, vocational, or some college, and college graduate or postgraduate); annual household income (<US $20,000, US $20,000-US $49,999, US $50,000-US $74,999, and ≥US $75,000); and geographical area. Geographical area was categorized as urban or rural based on the US Department of Agriculture Economic Research Service continuum codes.

Health-related variables included BMI (normal: 18.5-24.9; overweight: 25-29.9; and obese: ≥30), perceived health status (*poor* or *fair*, *good*, *very good*, and *excellent*), health self-efficacy (*not confident at all* or *a little confident*, *somewhat confident*, *very confident*, and *completely confident*), having one or more chronic conditions (diabetes, high blood pressure, or a heart condition), and being a health information seeker. Perceived health status was evaluated with a single item, *In general, would you say your health is…* This single item includes a 5-point Likert response from *poor* to *excellent*. Health self-efficacy was measured with the item, *Overall, how confident are you about your ability to take good care of your health?* Response options were also on a 5-point Likert scale, ranging from *completely confident* to *not confident at all*. Having one or more chronic conditions was reported using a checklist that followed the question, *Has a doctor or other health professional ever told you that you had any of the following medical conditions?* In this study, respondents were classified as having a chronic condition if they selected *yes* to *diabetes or high blood sugar*, *high blood pressure or hypertension*, or *a heart condition such as heart attack, angina, or congestive heart failure*. These chronic conditions were included because self-management using digital medical devices (eg, glucometer) is common. The item, *Have you ever looked for information about health or medical topics from any source?* (*yes* or *no*), was used to evaluate being a health information seeker.

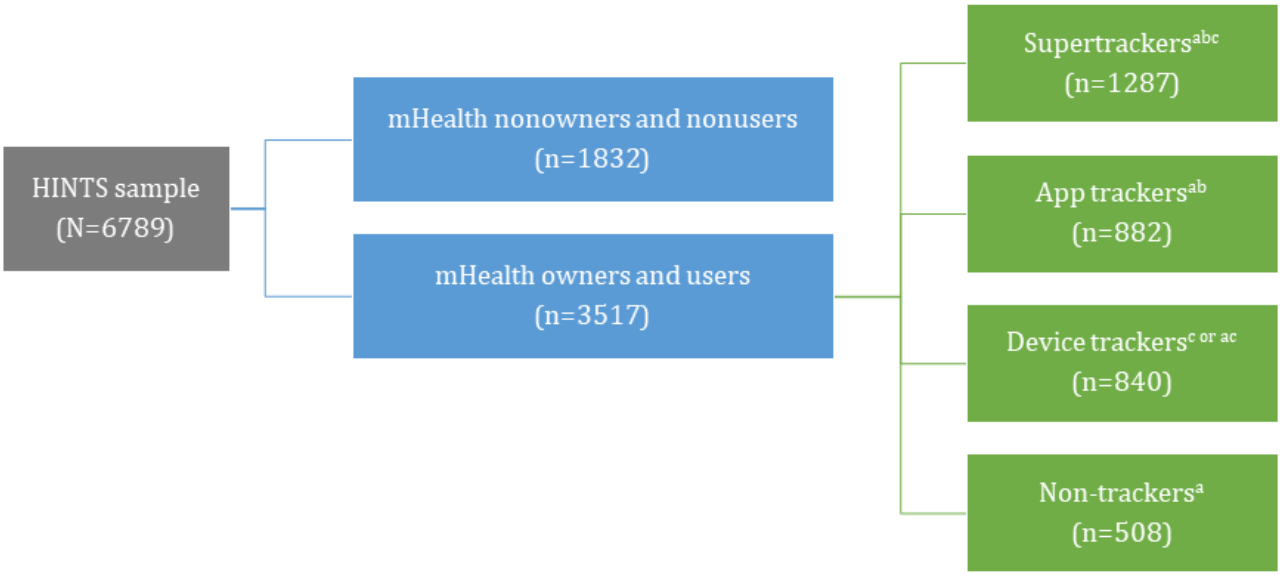
Mobile Health Technologies Owned and Used for Health or Behavioral Tracking and Goal Setting

Respondents with a smartphone or tablet computer were asked, *On your tablet or smartphone, do you have any apps related to health and wellness?* (*yes*, *no*, or *don't know*). For logistic regression analyses, this response was dichotomized to yes vs no and don't know. *Don't know* responses were collapsed with *no* responses owing to a low response frequency for *don't know* (n=285) and because not knowing whether one has health apps suggests nonuse of those apps for health or behavioral tracking. Respondents were also asked, *Has your tablet or smartphone helped you track progress on a health-related goal such as quitting smoking, losing weight, or increasing physical activity?* (*yes* or *no*). All respondents were asked, *Other than a tablet or smartphone, have you used an electronic device to monitor or track your health within the last 12 months? Examples include Fitbit, blood glucose meters, and blood pressure monitors*. (*yes* or *no*).

Operationalizing Pattern of Mobile Health Use for Health and Behavioral Tracking

Figure 1 represents our operationalization of each mutually exclusive pattern of mHealth use for health and behavioral tracking. The combined HINTS sample was divided into mHealth owners and users and nonowners and nonusers. mHealth nonowners and nonusers were defined as those who reported that they do not have or do not know if they have health apps on their smartphone or tablet, do not use a smartphone or tablet to track health-related goals, and do not use other devices to track health or behaviors (eg, Fitbit, glucose meter, and blood pressure monitor). mHealth owners and users were divided into 4 categories based on distinctive ways mHealth technologies are owned or used to track health or behaviors and goals. Supertrackers were defined as those who reported using a smartphone or tablet to track health-related goals and use other devices, such as a fitness tracker or digital medical device, to track health or behaviors. App trackers were defined as those who reported using a smartphone or tablet to track health-related goals but do not use other devices to track health or behaviors. Device trackers were defined as those who reported using devices other than a smartphone or tablet to track health or behaviors but do not use a smartphone or tablet to track health-related goals. Finally, nontrackers were defined as those who only reported having smartphone or tablet health apps but do not use them to track health-related goals and do not use other devices to track health or behaviors.

Figure 1. Operationalizing pattern of mobile health use for health and behavioral tracking. a. On your tablet or smartphone, do you have any "apps" related to health and wellness? Yes. b. Has your tablet or smartphone helped you track progress on a health-related goal such as quitting smoking, losing weight, or increasing physical activity? Yes. c. Other than a tablet or smartphone, have you used an electronic device to monitor or track your health within the last 12 months? Examples include Fitbit, blood glucose meters, and blood pressure monitors. Yes. HINTS: Health Information National Trends Survey.



Statistical Analysis

Analyses of merged HINTS 5, Cycle 1 and HINTS 5, Cycle 2 data were conducted using SAS version 9.4 (SAS Institute, Cary, NC). Descriptive statistics (frequencies, weighted percentages, and chi-square statistics) were used to examine patterns of mHealth use for health and behavioral tracking (Aim 1). Respondents with and without a smartphone or tablet were included in the mHealth nonowners and nonusers category; not having, not tracking, and not monitoring health with a mobile device was the primary criterion for inclusion. Similar to mHealth nonowners and nonusers, having a smartphone or tablet was not a criterion for inclusion in the device tracker category.

In all, 2 logistic regression models were constructed to characterize and compare US population subgroups by pattern of mHealth use for health and behavioral tracking. To address Aim 2, binomial logistic regression was performed to test the odds that respondent characteristics independently predict being in the mHealth nonowners and nonusers category, while holding the other respondent characteristics constant (reference category=all mHealth owners and users). To address Aim 3, a multinomial logistic regression model using the generalized logit was constructed to determine the predicted probability that respondent characteristics independently predict mHealth owners' and users' pattern of mHealth use for health and behavioral tracking, while controlling for the other respondent characteristics. Because a normative group of mHealth owners and users does not exist, supertrackers were chosen as the reference category for this second model as Aim 1 findings showed supertrackers made up the largest category of mHealth owners and users in the sample. Tests for significance for odds ratios and 95% CI were calculated at $P<.05$. Complete case analysis and listwise deletion were used for regression models.

Data weights were applied to provide representative US population estimates. A full-sample weight was used to calculate population-level point estimates and parameters. Replicate weights (calculated using the jackknife variance estimation method) were used to compute accurate standard errors.

Results

Sociodemographic and Health-Related Characteristics

The total combined HINTS sample was 6789. Weighted prevalence estimates were calculated for HINTS respondents' sociodemographic and health-related characteristics (see Multimedia Appendix 3). HINTS poststratification weightings adjust for inherent variations in the obtained sample by adjusting weighted totals to approximate known population characteristics. As in most health surveys, the HINTS raw sample tended to overaccumulate responses from females, older age individuals, people who are urban-dwelling, and those with more education and higher annual household income. In addition, white, non-Hispanic respondents comprised more than half of the sample. With respect to health-related characteristics, about half of the US adult population reported at least very good perceived health status, and most were at least very confident about their health self-efficacy. Most were health information seekers, overweight or obese, and had never been diagnosed with diabetes, hypertension, or a heart condition.

Mobile Health Technologies Owned and Used for Health or Behavioral Tracking and Goal Setting

Weighted prevalence estimates were also calculated for items pertaining to ownership and use of mHealth technologies for health or behavioral tracking and goal setting. Most respondents (85.22%) reported having a smartphone or tablet, close to half (46.83%) reported having a health app on one of these devices, and 43.23% reported that they use their smartphone or tablet to track progress on a health-related goal. Approximately one-third of respondents (34.66%) reported that they have a device other than a smartphone or tablet to track their health or behaviors.

Aim 1: Patterns of Mobile Health Use for Health and Behavioral Tracking

On the basis of our operationalization of pattern of mHealth use for health and behavioral tracking, mHealth nonowners and nonusers made up about one-third of the sample (34.72%). Most mHealth nonowners and nonusers were owners of a smartphone or tablet (95.15%). Nontrackers, those who reported having health apps on their smartphone or tablet but do not use them to track a health-related goal and do not use other devices to track health or behaviors, comprised the smallest percentage of the sample (8.8%).

Among respondents who reported owning or using mHealth technologies for health or behavioral tracking, supertrackers were the most prevalent (24.05%), followed by app trackers (19.5%) and device trackers (12.9%). Most device trackers reported having a smartphone or tablet (97.6%), and approximately one-third reported that they have a health app on one of those devices (35.0%). Table 1 describes the sociodemographic and health-related characteristics of respondents by pattern of mHealth use for health and behavioral tracking.

Table 1. Weighted population estimates for sociodemographic and health-related characteristics by pattern of mobile health use for health and behavioral tracking; HINTS 5, Cycle 1, 2017 and Cycle 2, 2018 (N=6789).

Characteristics	Nonowners and nonusers (n=1832; 34.72%)	Nontrackers (n=508; 8.8%)	Device trackers (n=840; 12.9%)	App trackers (n=882; 19.5%)	Supertrackers (n=1287; 24.05%)	Chi-square (*df*)	*P* value
Gender, n (%)						21.6 (4)	<.001
Male	733 (51.90)	192 (46.4)	381 (58.0)	298 (46.2)	445 (41.93)	N/A[a]	N/A
Female	955 (48.10)	290 (53.6)	396 (42.0)	542 (53.8)	776 (58.07)	N/A	N/A
Age (years), n (%)						222.5 (12)	<.001
18-34	180 (21.25)	61 (22.6)	42 (12.0)	244 (42.1)	216 (27.56)	N/A	N/A
35-49	355 (28.47)	117 (29.4)	96 (22.1)	267 (33.1)	390 (33.67)	N/A	N/A
50-64	645 (33.36)	183 (33.5)	289 (37.9)	255 (19.9)	458 (30.76)	N/A	N/A
≥65	583 (16.92)	133 (14.5)	388 (28.0)	101 (4.9)	203 (8.01)	N/A	N/A
Race and ethnicity, n (%)						29.3 (12)	.004
White, non-Hispanic	1,046 (64.50)	315 (70.7)	538 (75.0)	487 (60.9)	756 (63.92)	N/A	N/A
Black, non-Hispanic	204 (9.91)	60 (10.5)	87 (6.9)	130 (10.5)	182 (11.60)	N/A	N/A
Hispanic	272 (17.72)	52 (12.9)	83 (11.0)	148 (19.1)	155 (13.69)	N/A	N/A
Other	142 (7.87)	45 (5.9)	62 (7.1)	72 (9.5)	123 (10.79)	N/A	N/A
Education, n (%)						109.9 (8)	<.001
High school graduate or less	514 (35.42)	84 (28.5)	170 (26.3)	150 (21.2)	148 (16.37)	N/A	N/A
Technical, vocational, or some college	566 (38.31)	150 (30.9)	284 (43.2)	259 (38.5)	337 (36.84)	N/A	N/A
College graduate or postgraduate	705 (26.27)	266 (40.6)	369 (30.5)	466 (40.3)	791 (46.79)	N/A	N/A
Income in US $, n (%)						106.5 (12)	<.001
<20,000	360 (19.04)	45 (9.4)	109 (12.0)	125 (14.9)	104 (10.23)	N/A	N/A
20,000-49,999	530 (29.04)	111 (18.6)	242 (27.4)	223 (27.7)	214 (14.91)	N/A	N/A
50,000-74,999	342 (20.76)	100 (19.1)	159 (19.4)	160 (16.9)	243 (19.93)	N/A	N/A
≥75,000	566 (31.16)	245 (52.9)	316 (41.2)	367 (40.5)	712 (54.93)	N/A	N/A
Geographical area, n (%)						15.4 (4)	.004
Urban	1569 (85.83)	460 (90.2)	709 (82.4)	782 (87.8)	1,170 (90.39)	N/A	N/A
Rural	263 (14.17)	48 (9.8)	131 (17.6)	100 (12.2)	117 (9.61)	N/A	N/A
Perceived health status, n (%)						47.4 (12)	<.001
Poor or fair	281 (14.70)	58 (8.2)	179 (21.5)	102 (11.7)	141 (9.89)	N/A	N/A
Good	643 (35.14)	170 (36.5)	315 (38.8)	257 (31.2)	399 (30.42)	N/A	N/A
Very good	672 (36.92)	206 (41.9)	267 (30.7)	371 (37.4)	551 (45.23)	N/A	N/A
Excellent	219 (13.24)	70 (13.4)	70 (9.0)	145 (19.7)	189 (14.46)	N/A	N/A
Health self-efficacy, n (%)						19.4 (12)	.08
A little confident or not confident at all	89 (5.11)	15 (2.7)	51 (8.3)	22 (3.2)	35 (3.50)	N/A	N/A
Somewhat confident	442 (26.83)	112 (25.5)	221 (26.3)	174 (20.1)	276 (23.31)	N/A	N/A
Very confident	845 (44.51)	237 (46.5)	383 (45.2)	412 (44.9)	634 (48.53)	N/A	N/A
Completely confident	440 (23.55)	138 (25.3)	178 (20.2)	268 (31.8)	337 (24.66)	N/A	N/A
Health information seeker, n (%)						66.15 (4)	<.001

Characteristics	Nonowners and nonusers (n=1832; 34.72%)	Nontrackers (n=508; 8.8%)	Device trackers (n=840; 12.9%)	App trackers (n=882; 19.5%)	Supertrackers (n=1287; 24.05%)	Chi-square (*df*)	*P* value
Yes	1,369 (76.76)	438 (84.2)	712 (82.8)	757 (85.1)	1,191 (92.43)	N/A	N/A
No	439 (23.24)	59 (15.8)	118 (17.2)	118 (14.9)	84 (7.57)	N/A	N/A
BMI, n (%)						19.5 (8)	.01
Normal (18.5-24.9)	565 (33.11)	195 (41.1)	202 (26.6)	286 (35.1)	344 (28.19)	N/A	N/A
Overweight (25-29.9)	610 (33.61)	152 (30.2)	298 (31.9)	301 (34.6)	451 (36.19)	N/A	N/A
Obese (≥30)	561 (33.28)	145 (28.7)	304 (41.5)	269 (30.3)	452 (35.62)	N/A	N/A
One or more chronic conditions, n (%)						104.0 (4)	<.001
Yes	848 (35.71)	213 (34.6)	609 (65.6)	277 (26.5)	594 (39.09)	N/A	N/A
No	984 (64.29)	295 (65.4)	231 (34.4)	605 (73.5)	693 (60.91)	N/A	N/A

[a]N/A: not applicable.

Aim 2: Characteristics Associated With Being in the Mobile Health Nonowners and Nonusers Category

Compared with mHealth owners and users (nontrackers, app trackers, device trackers, and supertrackers), mHealth nonowners and nonusers had higher odds of being male (OR 1.38, 95% CI 1.08-1.77) and 35 years old or older, with the odds of being in the mHealth nonowners and nonusers category increasing with advancing age (35-49 years old: OR 1.55, 95% CI 1.01-2.36; 50-64 years old: OR 2.01, 95% CI 1.30-3.11; 65+ years old: OR 2.77, 95% CI 1.86-4.13). mHealth nonowners and nonusers also had higher odds of reporting an annual household income below US $75,000 (<US $20,000: OR 2.06, 95% CI 1.34-3.16; US $20,000-US $49,999: OR 1.94, 95% CI 1.33-2.83; US $50,000-US $74,999: OR 1.67, 95% CI 1.14-2.46).

In addition, mHealth nonowners and nonusers had higher odds of not being a college graduate (high school graduate or less: OR 1.88, 95% CI 1.31-2.69; technical, vocational, or some college: OR 1.36, 95% CI 1.02-1.83) and not being health information seekers (OR 1.53, 95% CI 1.15-2.04). However, they had lower odds of reporting a chronic condition (OR 0.57, 95% CI 0.45-0.72) compared with mHealth owners and users. There were nonsignificant differences for race and ethnicity ($F_{3,98}$=0.11; P=.95), geographical area ($F_{1,98}$=0.39; P=.54), perceived health status ($F_{3,98}$=0.31; P=.82), health self-efficacy ($F_{3,98}$=0.92; P=.43), and weight status (BMI, $F_{1,98}$=1.77; P=.18) when controlling all other variables (see Table 2).

Table 2. Sociodemographic and health-related characteristics significantly associated with being in the mobile health nonowners and nonusers category (n=1468). Reference category=all mobile health owners and users.

Characteristic	Odds ratio (95% CI)	*P* value
Gender		
Female	Reference	N/A[a]
Male	1.38 (1.08-1.77)	.01
Age (years)		
18-34	Reference	N/A
35-49	1.55 (1.01-2.36)	.04
50-64	2.01 (1.30-3.11)	.002
≥65	2.77 (1.86-4.13)	<.001
Education		
College graduate or postgraduate	Reference	N/A
Technical, vocational, or some college	1.36 (1.02-1.83)	.04
High school graduate or less	1.88 (1.31-2.69)	.001
Income in US $		
≥75,000	Reference	N/A
50,000-74,999	1.67 (1.14-2.46)	.009
20,000-49,999	1.94 (1.33-2.83)	<.001
<20,000	2.06 (1.34-3.16)	.001
Health information seeker		
Yes	Reference	N/A
No	1.53 (1.15-2.04)	.004
One or more chronic conditions		
No	Reference	N/A
Yes	0.57[b] (0.45-0.72)	<.001

[a]N/A: not applicable.

[b]Indicates a negative association.

Aim 3: Characteristics Associated With Mobile Health Owners' and Users' Pattern of Mobile Health Use for Health and Behavioral Tracking

Multinomial logistic regression analysis revealed several sociodemographic and health-related characteristics associated with pattern of mHealth use for health and behavioral tracking among mHealth owners and users (reference category=supertrackers). As shown in Table 3, there were significant differences in terms of gender, age, race and ethnicity, income, having a chronic condition, weight status, and being a health information seeker. There were, however, nonsignificant differences for pattern of mHealth use by geographical area ($F_{3,98}$=1.37; *P*=.29), perceived health status ($F_{9,98}$=1.33; *P*=.23), health self-efficacy ($F_{9,98}$=0.91; *P*=.52), and education ($F_{6,98}$=1.47; *P*=.20) when controlling for all other variables.

Compared with supertrackers, the largest category of mHealth owners and users in this study, app trackers had higher odds of being male (OR 1.46, 95% CI 1.11-1.94) and lower odds of being 50 years old or older (50-64 years old: OR 0.63, 95% CI 0.41-0.96; 65+ years old: OR 0.50, 95% CI 0.29-0.88). App trackers had about 2 times the odds of reporting an annual household income in the US $20,000-US $49,999 range than US $75,000 or more (OR 2.32, 95% CI 1.61-3.33). Finally, app trackers had lower odds of reporting that they have a chronic condition than supertrackers (OR 0.62, 95% CI 0.42-0.92).

Device trackers also had higher odds of being male (OR 2.22, 95% CI 1.55-3.19) compared with supertrackers. With respect to age, device trackers had nearly 3 times the odds of being 50-64 years old (OR 2.83, 95% CI 1.52-5.30) and about 6 times the odds of being 65 years old or older (OR 6.28, 95% CI 3.35-11.79). Device trackers had about 2 times the odds of reporting an annual household income in the US $20,000-US $49,999 range than US $75,000 or more when compared with supertrackers (OR 2.31, 95% CI 1.36-3.91) and about 3 times the odds of not being health information seekers (OR 2.98, 95% CI 1.66-5.33). Device trackers had lower odds of identifying as non-Hispanic black (OR 0.48, 95% CI 0.31-0.74) and other race and ethnicity (OR 0.54, 95% CI 0.30-0.99). They had higher odds of reporting a chronic condition (OR 1.69, 95% CI

1.14-2.49) but lower odds of being overweight (OR 0.58, 95% CI 0.38-0.88) or obese (OR 0.55, 95% CI 0.35-0.86) than supertrackers.

Finally, compared with supertrackers, nontrackers had more than twice the odds of being 65 years old or older (OR 2.74, 95% CI 1.47-5.11). They also had more than 2 times the odds of not being health information seekers (OR 2.37, 95% CI 1.16-4.86). Nontrackers had lower odds of being in the other race and ethnicity category (OR 0.51, 95% CI 0.27-0.96) than supertrackers. In addition, they had lower odds of being overweight (OR 0.47, 95% CI 0.29-0.76) or obese (OR 0.43, 95% CI 0.28-0.67).

Table 3. Characteristics significantly associated with being a mobile health (mHealth) owner and user by pattern of mHealth use for health and behavioral tracking. Reference category=Supertrackers (n=1122).

Characteristic	App trackers (n=773), OR (95% CI)	*P* value	Device trackers (n=685), OR (95% CI)	*P* value	Nontrackers (n=428), OR (95% CI)	*P* value
Gender						
Female (ref)	N/A[a]	N/A	N/A	N/A	N/A	N/A
Male	1.46 (1.11-1.94)	.008	2.22 (1.55-3.19)	<.001	1.47 (0.96-2.25)	.07
Age (years)						
18-34 (ref)	N/A	N/A	N/A	N/A	N/A	N/A
35-49	0.80 [b] (0.54-1.19)	.27	1.41 (0.70-2.85)	.34	1.29 (0.70-2.38)	.42
50-64	0.63 [b] (0.41-0.96)	.03	2.83 (1.52-5.30)	.001	1.65 (0.84-3.24)	.14
>65	0.50[b] (0.29-0.88)	.02	6.28 (3.35-11.79)	<.001	2.74 (1.47-5.11)	.002
Race and ethnicity						
White, non-Hispanic (ref)	N/A	N/A	N/A	N/A	N/A	N/A
Black, non-Hispanic	1.02 (0.65-1.62)	.93	0.48[b] (0.31-0.74)	.001	0.95[b] (0.51-1.75)	.86
Hispanic	1.23 (0.80-1.88)	.35	0.75[b] (0.44-1.29)	.29	0.82[b] (0.42-1.60)	.56
Other	0.92[b] (0.47-1.81)	.81	0.54[b] (0.30-0.99)	.046	0.51[b] (0.27-0.96)	.04
Income in US $						
>75,000 (ref)	N/A	N/A	N/A	N/A	N/A	N/A
50,000-74,999	1.22 (0.78-1.92)	.39	1.31 (0.81-2.12)	.27	1.10 (0.66-1.82)	.71
20,000-49,999	2.32 (1.61-3.33)	<.001	2.31 (1.36-3.91)	.002	1.25 (0.71-2.20)	.44
<20,000	1.76 (0.83-3.73)	.14	1.33 (0.67-2.65)	.41	0.87[b] (0.33-2.31)	.78
Health information seeker						
Yes (ref)	N/A	N/A	N/A	N/A	N/A	N/A
No	1.70 (0.86-3.35)	.12	2.98 (1.66-5.33)	<.001	2.37 (1.16-4.86)	.02
BMI (kg/m^2)						
Normal (18.5-24.9; ref)	N/A	N/A	N/A	N/A	N/A	N/A
Overweight (25-29.9)	0.87[b] (0.57-1.33)	.51	0.58[b] (0.38-0.88)	.01	0.47[b] (0.29-0.76)	.002
Obese (≥30)	0.71[b] (0.45-1.14)	.15	0.55[b] (0.35-0.86)	.009	0.43[b] (0.28-0.67)	<.001
One or more chronic conditions						
No (ref)	N/A	N/A	N/A	N/A	N/A	N/A
Yes	0.62[b] (0.42-0.92)	.02	1.69 (1.14-2.49)	.009	0.67[b] (0.44-1.02)	.06

[a]N/A: not applicable.

[b]Indicates a negative association.

Discussion

Principal Findings

The aims of this study were to examine patterns of mHealth use for health and behavioral tracking in the US population and to characterize the population according to those patterns. We found that those who do not own or use mHealth technologies for health or behavioral tracking make up a relatively large proportion of the population (about one-third). Among owners and users of mHealth technologies, being a supertracker—using multiple mHealth technologies for health or behavioral tracking and goal setting—is the most common pattern of mHealth use. Our study also confirms and extends what other studies of nationally representative samples (eg, Canada) have discovered [7]; that is, there are substantive differences among mHealth technology owners and users in terms of sociodemographics and health-related factors. The sociodemographic and health-related characteristics of supertrackers and device trackers were the most distinctive, with several characteristics of device trackers—those who track health or behaviors but not health goals—paralleling those of mHealth nonowners and nonusers. Our findings suggest that the distinctive characteristics of supertrackers and device trackers, in particular, can be used to help target and tailor future mHealth interventions.

Characteristics of Nonowners and Nonusers of Mobile Health Technologies for Health and Behavioral Tracking

People who do not own or use mHealth technologies for health and behavioral tracking make up about one-third of the US population; regardless of the high percentage of Americans who own a smartphone or tablet, ownership does not necessarily mean that these devices are being used to track health, behaviors, or goals. We found that mHealth nonowners and nonusers tend to be male, older age, and report lower education and income levels than owners and users of mHealth technologies for health or behavioral tracking, which are findings supported by other studies focused on the US population's mHealth technology usage [5,24]. In addition, our findings revealed a relationship between being in the mHealth nonowners and nonusers category and not seeking health information in general.

mHealth nonowners and nonusers also tended not to have a chronic condition, specifically diabetes, hypertension, or a heart condition, when compared with owners and users of mHealth technologies for health or behavioral tracking. These results suggest that having a chronic condition is a potential reason or motivator for owning and using mHealth technologies. Although a study of smartphone owners found that those with chronic conditions are no more likely to use health apps than people without a chronic condition [10], our analyses also accounted for respondents' use of other devices, such as Fitbits, glucometers, and blood pressure monitors. Thus, our analyses appear to have detected distinctive differences in use of mHealth technologies between those with and without a chronic condition because these diverse types of technology were included in our measurement.

Although the link between older age and being a nonowner or nonuser of mHealth technologies is consistently reported across studies, we recommend continued population-level research of mHealth nonowners and nonusers because the proportion of digital natives will grow as the population ages. In turn, this may increase the proportion of the population that trusts collecting personal health information on digital devices [26]. Perceived utility of mHealth technologies for health or behavioral tracking may also increase among nonowners and nonusers and former owners and users as mHealth technologies advance in their functionality, especially if approaches such as ensuring sociocultural relevancy [27] and person-centered design [28,29] are considered throughout mHealth technology development.

In addition, our findings related to mHealth nonowners and nonusers suggest that clinical and public health practitioners could consider alternatives to mHealth interventions to track and promote health behavior change, especially among less educated older men. However, they also call to mind the importance of ensuring that not owning or using mHealth technologies for health or behavioral tracking is not due to barriers that can be addressed, such as digital health literacy [28,30-32]. Researchers and organizations are addressing some of these issues through programs such as RecycleHealth [31,33], American Association of Retired Persons Tek workshops [34,35], and The Wellness Group [36], with measurable beneficial effects on health and behavioral outcomes [31,36].

Characteristics of Mobile Health Owners and Users by Pattern of Mobile Health Use for Health and Behavioral Tracking

In this study, we also discovered differences in the characteristics of mHealth owners and users by comparing nontrackers, app trackers, and device trackers with the largest group of mHealth owners and users, supertrackers. Supertrackers, those who use multiple devices for health or behavioral tracking and goal setting, make up approximately one-quarter of the US population. Our findings demonstrate that supertrackers, conceivably the most intrinsically interested in mHealth technologies, are younger than nontrackers and device trackers, who have in common that they do not use mHealth technologies to track a health-related goal. Supertrackers also tended to be female when compared with their tracker counterparts, device trackers and app trackers. This finding related to gender is consistent with previous studies of people who download smartphone health apps [24] and users of wearable activity monitors, such as Fitbits [5].

With respect to socioeconomic factors, when compared with supertrackers, we found that being an app tracker or being a device tracker, respectively, was associated with reporting an annual household income in the US $20,000-US $49,999 range vs US $75,000 or more. Paré et al [7] found in a Canadian sample that reporting the highest annual income level, greater than Can $80,000 (US $59968.80), was associated with being a *digital self-tracker*, someone who uses health apps, wearables, or digital medical devices. By analyzing the US population's distinctive patterns of mHealth use, we add to the literature that multiuse of mHealth technologies for health or behavioral

tracking and goal setting is associated with higher reported annual household income. In addition, level of education was not significantly associated with mHealth owners and users' pattern of mHealth use for health and behavioral tracking when controlling all other factors. Other studies have found that higher level of education is associated with having or downloading smartphone health apps [2,24,37], digital self-tracking [7], and reporting current use of a wearable activity monitor [5]. Future studies should continue to examine relationships between mHealth owners and users' specific pattern of mHealth use, education, and income, as ease of use, pricing, and accessibility will continue to evolve with the advancement of mHealth technologies.

Although we did not find that racial or ethnic identity was associated with being an mHealth owner or user vs being in the mHealth nonowners and nonusers category (see Aim 2 results), race and ethnicity was associated with mHealth owners and users' pattern of mHealth use for health and behavioral tracking. For example, non-Hispanic blacks and respondents in the other race and ethnicity category had lower odds of being device trackers than supertrackers. Future studies should explore the relationship between sociocultural factors that foster or challenge interest and motivation to use mHealth technologies and the perceived sociocultural relevancy of different types of mHealth technologies for health or behavioral tracking and goal setting [27].

We also discovered several health-related characteristics associated with mHealth owners and users' pattern of mHealth use for health and behavioral tracking. For example, we found that supertrackers tended to be health information seekers when compared with mHealth owners and users who do not use devices to track health goals (ie, nontrackers and device trackers). One explanation for this finding is that supertrackers may be relatively more attentive, interested, and curious about personal health information (evidenced by their goal-setting behavior), including collecting and recording it on digital devices [30,38]. Health information–seeking behavior and goal setting appear to be distinguishing characteristics among the US population's mHealth technology users.

Unlike the results of our analysis of mHealth nonowners and nonusers vs mHealth owners and users, we found a significant relationship between mHealth owners and users' pattern of mHealth use for health and behavioral tracking and weight status. Specifically, supertrackers had greater odds of being overweight or obese compared with nontrackers and device trackers, respectively. Krebs and Duncan [2] found that smartphone owners who download health apps are more likely to be overweight or obese, and Byuhan et al [37] also report greater odds of using health apps to track a health-related goal among obese compared with underweight individuals. Therefore, we add to the findings of these authors that being overweight or obese is associated with multiuse of mHealth technologies for health or behavioral tracking and goal setting.

Having a chronic condition (diabetes, hypertension, or a heart condition) was also associated with mHealth owners and users' pattern of mHealth use for health and behavioral tracking; however, the direction of the relationship differed when comparing supertrackers with app trackers and with device trackers. Namely, supertrackers had greater odds of having a chronic condition than app trackers but lower odds of having a chronic condition than device trackers. This result may be explained by the wide range of devices other than a smartphone or tablet that respondents might use, ranging from fitness trackers to digital medical devices, such as glucometers. On the basis of the characteristics of device trackers—male, older age, non–health information seekers, with a chronic condition—the type of mHealth technologies used by this group may be digital medical devices over fitness trackers. We recommend that future iterations of HINTS include separate items to measure use of specific types of wearables and other portable digital devices because motivations for use and individual differences of mHealth technology users likely vary considerably by the specific type of nonsmartphone or nontablet mHealth technology used.

We also recommend that mHealth intervention researchers and health care providers who study or prescribe mHealth to individuals who fit the device tracker profile consider individual-level factors that might threaten continued engagement with mHealth technologies, such as relatively less intrinsic interest in digital self-tracking, potentially low digital health literacy, and data-entry burden [2,29,36]. However, our findings suggest opportunities, not only potential challenges. For example, although device trackers in our study reported that they do not use health apps to track health-related goals, about one-third reported downloading health apps onto their smartphone or tablet. These downloads may represent implementation intentions that clinical and public health practitioners might leverage by working with device trackers to set personal health goals within their health apps. Health goal-setting and reminders embedded in mHealth technologies that stimulate habit formation may lead to *guided mastery*, which is theorized to help people with implementation intentions act on their behavioral intentions [22].

Notably, geographical area (urban vs rural), perceived health, and health self-efficacy were not significantly associated with pattern of mHealth use for health and behavioral tracking. These results are inconsistent with studies showing that mHealth technology users are more likely to reside in urban over rural areas [37]. One explanation for our findings may be the rapid increase in smartphone ownership throughout the entire United States since the time other population studies were conducted [1]. Although better perceived health status [7,10,24,29,37] and greater health self-efficacy [37] have been associated with mHealth technology use in past studies, our results highlight that other individual-level factors may be stronger predictors when mHealth technology owners and users are compared by distinctive pattern of mHealth use for health and behavioral tracking.

Limitations

This study is the largest and most current nationally representative study of mHealth ownership and use in US adults across 2 cycles of NCI HINTS. However, there are also several limitations, including reliance on self-report and cross-sectional data. Because the survey data are cross-sectional, we cannot

infer causal relationships between variables (eg, being a supertracker leads to obesity or vice versa). In addition, we cannot discern whether respondents intended to engage, reengage, or disengage from mHealth technologies following their participation in HINTS.

Given the limitations of the dataset, we could neither determine the type of devices used beyond smartphones and tablets nor discern whether respondents used a combination of nonsmartphone or nontablet devices (eg, fitness tracker plus a digital medical device). We also could not distinguish their specific reasons for tracking health or behaviors (eg, blood glucose monitoring or number of steps walked) or the extent to which they engage with mHealth technologies (frequency of use, duration of use, etc). Moreover, we were unable to evaluate their experience of mHealth technology use (eg, usability and acceptability). However, findings from this study may be used to guide development of future research focused on improving users' experience of mHealth technologies for health or behavioral tracking because we were able to capture the characteristics of the US population by diverse patterns of mHealth use.

An additional limitation of this study is that we could not determine the specific health goals respondents set in their health apps. Goals may have varied from smoking cessation to weight loss to asthma control, for example. Although the aim of this study was not to make predictions based on the specific goals of mHealth technology use, we acknowledge that making such distinctions is important for behavioral interventions.

Conclusions

This study contributes to understanding of the US adult population's ownership, use (including multiuse), and nonuse of diverse types of mHealth technologies for health or behavioral tracking and goal setting. We discovered that characteristics, such as age, gender, being a health information seeker, and having a chronic condition, are associated with specific patterns of mHealth use. Researchers and clinical and public health practitioners can apply these findings to research design, practice, and health message development to better reach intended audiences and promote health behavior change.

Although mHealth technologies have the potential to broadly reach people and facilitate behavior change, our findings suggest that an appreciation for the diverse ways mHealth technologies are used (and not used) for health or behavioral tracking and goal setting should be considered when designing as well as interpreting the results of mHealth intervention studies. Future studies could build on this work through continued surveillance of patterns of mHealth use for health and behavioral tracking and the individual-level factors associated with those patterns. Research that keeps pace with mHealth technology development is needed to understand the contextual factors that help explain variation in population-level effects of mHealth technology use on health behavior.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Health Information National Trends Survey 5, Cycle 1 instrument.
[PDF File (Adobe PDF File), 2225 KB - jmir_v22i5e16299_app1.pdf]

Multimedia Appendix 2
Health Information National Trends Survey 5, Cycle 2 instrument.
[PDF File (Adobe PDF File), 2502 KB - jmir_v22i5e16299_app2.pdf]

Multimedia Appendix 3
Weighted Population Estimates for Sociodemographic and Health-Related Characteristics, HINTS 5, Cycle 1 and Cycle 2.
[PDF File (Adobe PDF File), 78 KB - jmir_v22i5e16299_app3.pdf]

References

1. Taylor K, Silver L. Pew Research Center. 2019 Feb 5. Smartphone Ownership is Growing Rapidly Around the World, but Not Always Equally URL: https://www.pewresearch.org/global/wp-content/uploads/sites/2/2019/02/Pew-Research-Center_Global-Technology-Use-2018_2019-02-05.pdf [accessed 2019-08-20]
2. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov 4;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
3. Dunn J, Runge R, Snyder M. Wearables and the medical revolution. Per Med 2018 Sep;15(5):429-448. [doi: 10.2217/pme-2018-0044] [Medline: 30259801]
4. Hitlin P. Pew Research Center. 2018 Sep 28. Internet, Social Media Use and Device Ownership in US Have Plateaued After Years of Growth URL: https://www.pewresearch.org/fact-tank/2018/09/28/internet-social-media-use-and-device-ownership-in-u-s-have-plateaued-after-years-of-growth/ [accessed 2019-08-20]

5. Omura JD, Carlson SA, Paul P, Watson KB, Fulton JE. National physical activity surveillance: users of wearable activity monitors as a potential data source. Prev Med Rep 2017 Mar;5:124-126 [FREE Full text] [doi: 10.1016/j.pmedr.2016.10.014] [Medline: 28101443]
6. Jiang S, Hong YA, Liu PL. Trends of online patient-provider communication among cancer survivors from 2008 to 2017: a digital divide perspective. J Cancer Surviv 2019 Apr;13(2):197-204. [doi: 10.1007/s11764-019-00742-4] [Medline: 30756225]
7. Paré G, Leaver C, Bourget C. Diffusion of the digital health self-tracking movement in Canada: results of a national survey. J Med Internet Res 2018 May 2;20(5):e177 [FREE Full text] [doi: 10.2196/jmir.9388] [Medline: 29720359]
8. Yingling LR, Mitchell V, Ayers CR, Peters-Lawrence M, Wallen GR, Brooks AT, et al. Adherence with physical activity monitoring wearable devices in a community-based population: observations from the Washington, D.C., Cardiovascular Health and Needs Assessment. Transl Behav Med 2017 Dec;7(4):719-730 [FREE Full text] [doi: 10.1007/s13142-016-0454-0] [Medline: 28097627]
9. Tuman M, Moyer A. Health intentions and behaviors of health app owners: a cross-sectional study. Psychol Health Med 2019 Aug;24(7):819-826. [doi: 10.1080/13548506.2019.1576911] [Medline: 30729803]
10. Robbins R, Krebs P, Jagannathan R, Jean-Louis G, Duncan DT. Health app use among US mobile phone users: analysis of trends by chronic disease status. JMIR Mhealth Uhealth 2017 Dec 19;5(12):e197 [FREE Full text] [doi: 10.2196/mhealth.7832] [Medline: 29258981]
11. Wang Y, Xue H, Huang Y, Huang L, Zhang D. A systematic review of application and effectiveness of mhealth interventions for obesity and diabetes treatment and self-management. Adv Nutr 2017 May;8(3):449-462 [FREE Full text] [doi: 10.3945/an.116.014100] [Medline: 28507010]
12. Brickwood KJ, Watson G, O'Brien J, Williams AD. Consumer-based wearable activity trackers increase physical activity participation: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019 Apr 12;7(4):e11819 [FREE Full text] [doi: 10.2196/11819] [Medline: 30977740]
13. Schoeppe S, Alley S, van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 7;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
14. Qudah B, Luetsch K. The influence of mobile health applications on patient - healthcare provider relationships: a systematic, narrative review. Patient Educ Couns 2019 Jun;102(6):1080-1089. [doi: 10.1016/j.pec.2019.01.021] [Medline: 30745178]
15. Buman MP, Epstein DR, Gutierrez M, Herb C, Hollingshead K, Huberty JL, et al. BeWell24: development and process evaluation of a smartphone 'app' to improve sleep, sedentary, and active behaviors in US Veterans with increased metabolic risk. Transl Behav Med 2016 Sep;6(3):438-448 [FREE Full text] [doi: 10.1007/s13142-015-0359-3] [Medline: 27528532]
16. Carpenter DM, Geryk LL, Sage A, Arrindell C, Sleath BL. Exploring the theoretical pathways through which asthma app features can promote adolescent self-management. Transl Behav Med 2016 Dec;6(4):509-518 [FREE Full text] [doi: 10.1007/s13142-016-0402-z] [Medline: 27118115]
17. Desveaux L, Shaw J, Saragosa M, Soobiah C, Marani H, Hensel J, et al. A mobile app to improve self-management of individuals with type 2 diabetes: qualitative realist evaluation. J Med Internet Res 2018 Mar 16;20(3):e81 [FREE Full text] [doi: 10.2196/jmir.8712] [Medline: 29549070]
18. Gordon JS, Armin J, Hingle MD, Giacobbi P, Cunningham JK, Johnson T, et al. Development and evaluation of the See Me Smoke-Free multi-behavioral mHealth app for women smokers. Transl Behav Med 2017 Jun;7(2):172-184 [FREE Full text] [doi: 10.1007/s13142-017-0463-7] [Medline: 28155107]
19. Hesse BW, Beckjord E, Ahern DK. Role of technology in behavior change to expand reach and impact on public health. In: Hilliard ME, Riekert KA, Ockene JK, Pbert L, editors. The Handbook of Health Behavior Change. Fifth Edition. New York: Springer; 2018:525-544.
20. Locke EA, Latham GP. Building a practically useful theory of goal setting and task motivation. A 35-year odyssey. Am Psychol 2002 Sep;57(9):705-717. [doi: 10.1037//0003-066x.57.9.705] [Medline: 12237980]
21. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
22. Pirolli P, Youngblood GM, Du H, Konrad A, Nelson L, Springer A. Scaffolding the mastery of healthy behaviors with Fittle+ systems: Evidence-based interventions and theory. Hum–Comput Interact 2018 Oct 15:1-34. [doi: 10.1080/07370024.2018.1512414]
23. Samdal GB, Eide GE, Barth T, Williams G, Meland E. Effective behaviour change techniques for physical activity and healthy eating in overweight and obese adults; systematic review and meta-regression analyses. Int J Behav Nutr Phys Act 2017 Mar 28;14(1):42 [FREE Full text] [doi: 10.1186/s12966-017-0494-y] [Medline: 28351367]
24. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J Med Internet Res 2017 Apr 19;19(4):e125 [FREE Full text] [doi: 10.2196/jmir.5604] [Medline: 28428170]

25. Rutten LJ, Davis T, Beckjord EB, Blake K, Moser RP, Hesse BW. Picking up the pace: changes in method and frame for the health information national trends survey (2011-2014). J Health Commun 2012;17(8):979-989 [FREE Full text] [doi: 10.1080/10810730.2012.700998] [Medline: 23020763]
26. Serrano KJ, Yu M, Riley WT, Patel V, Hughes P, Marchesini K, et al. Willingness to exchange health information via mobile devices: Findings from a population-based survey. Ann Fam Med 2016;14(1):34-40 [FREE Full text] [doi: 10.1370/afm.1888] [Medline: 26755781]
27. Abroms LC, Allegrante JP, Auld ME, Gold RS, Riley WT, Smyser J. Toward a common agenda for the public and private sectors to advance digital health communication. Am J Public Health 2019 Feb;109(2):221-223. [doi: 10.2105/AJPH.2018.304806] [Medline: 30649936]
28. Kreps GL. The relevance of health literacy to mHealth. Inform Serv Use 2017;37(2):123-130. [doi: 10.3233/ISU-170828]
29. Simblett S, Greer B, Matcham F, Curtis H, Polhemus A, Ferrão J, et al. Barriers to and facilitators of engagement with remote measurement technology for managing health: systematic review and content analysis of findings. J Med Internet Res 2018 Jul 12;20(7):e10480 [FREE Full text] [doi: 10.2196/10480] [Medline: 30001997]
30. Cho J, Park D, Lee HE. Cognitive factors of using health apps: systematic analysis of relationships among health consciousness, health information orientation, eHealth literacy, and health app use efficacy. J Med Internet Res 2014 May 9;16(5):e125 [FREE Full text] [doi: 10.2196/jmir.3283] [Medline: 24824062]
31. Gualtieri L, Phillips J, Rosenbluth S, Synoracki S. Digital literacy: A barrier to adoption of connected health technologies in older adults. iproc 2018;4(2):e11803. [doi: 10.2196/11803]
32. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015 Mar 17;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]
33. RecycleHealth. URL: http://www.recyclehealth.com [accessed 2019-08-20]
34. American Association of Retired Persons. Learn@50+ URL: https://learn.aarp.org/ [accessed 2019-08-20]
35. Milani RV, Bober RM, Lavie CJ. The role of technology in chronic disease care. Prog Cardiovasc Dis 2016;58(6):579-583. [doi: 10.1016/j.pcad.2016.01.001] [Medline: 26772623]
36. Gualtieri L, Rosenbluth S, Phillips J. Can a free wearable activity tracker change behavior? The impact of trackers on adults in a physician-led wellness group. JMIR Res Protoc 2016 Nov 30;5(4):e237 [FREE Full text] [doi: 10.2196/resprot.6534] [Medline: 27903490]
37. Bhuyan SS, Lu N, Chandak A, Kim H, Wyant D, Bhatt J, et al. Use of mobile health applications for health-seeking behavior among US adults. J Med Syst 2016 Jun;40(6):153. [doi: 10.1007/s10916-016-0492-7] [Medline: 27147516]
38. O'Brien HL, Toms EG. What is user engagement? A conceptual framework for defining user engagement with technology. J Am Soc Inf Sci Technol 2008;59(6):938-955. [doi: 10.1002/asi.20801]

Abbreviations

HINTS: Health Information National Trends Survey
mHealth: mobile health
NCI: National Cancer Institute
OR: odds ratio

Edited by G Eysenbach; submitted 17.09.19; peer-reviewed by B Hesse, S Leveille, A Eskandar; comments to author 04.11.19; revised version received 23.12.19; accepted 03.02.20; published 14.05.20.

Please cite as:
Rising CJ, Jensen RE, Moser RP, Oh A
Characterizing the US Population by Patterns of Mobile Health Use for Health and Behavioral Tracking: Analysis of the National Cancer Institute's Health Information National Trends Survey Data
J Med Internet Res 2020;22(5):e16299
URL: https://www.jmir.org/2020/5/e16299
doi:10.2196/16299
PMID:32406865

Original Paper

Autoencoder as a New Method for Maintaining Data Privacy While Analyzing Videos of Patients With Motor Dysfunction: Proof-of-Concept Study

Marcus D'Souza[1*], MD; Caspar E P Van Munster[2*], MD; Jonas F Dorn[3], PhD; Alexis Dorier[3], MSc; Christian P Kamm[4,5], MD; Saskia Steinheimer[5], MD; Frank Dahlke[3], MD, PhD; Bernard M J Uitdehaag[2], MD, PhD; Ludwig Kappos[1], MD; Matthew Johnson[6], PhD

[1]Neurologic Clinic and Policlinic, Departments of Medicine, Biomedicine and Clinical Research, University Hospital Basel and University of Basel, Basel, Switzerland

[2]Department of Neurology, Multiple Sclerosis Center Amsterdam, Amsterdam University Medical Centers, Amsterdam, Netherlands

[3]Novartis Pharma AG, Basel, Switzerland

[4]Neurocenter, Luzerner Kantonsspital, Luzern, Switzerland

[5]Department of Neurology, Inselspital, University of Bern, Bern, Switzerland

[6]Microsoft Research, Cambridge, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Marcus D'Souza, MD
Neurologic Clinic and Policlinic
Departments of Medicine, Biomedicine and Clinical Research
University Hospital Basel and University of Basel
Petersgraben 4
Basel, 4031
Switzerland
Phone: 41 61 265 4151
Email: marcus.dsouza@usb.ch

Abstract

Background: In chronic neurological diseases, especially in multiple sclerosis (MS), clinical assessment of motor dysfunction is crucial to monitor the disease in patients. Traditional scales are not sensitive enough to detect slight changes. Video recordings of patient performance are more accurate and increase the reliability of severity ratings. When these recordings are automated, quantitative disability assessments by machine learning algorithms can be created. Creation of these algorithms involves non–health care professionals, which is a challenge for maintaining data privacy. However, autoencoders can address this issue.

Objective: The aim of this proof-of-concept study was to test whether coded frame vectors of autoencoders contain relevant information for analyzing videos of the motor performance of patients with MS.

Methods: In this study, 20 pre-rated videos of patients performing the finger-to-nose test were recorded. An autoencoder created encoded frame vectors from the original videos and decoded the videos again. The original and decoded videos were shown to 10 neurologists at an academic MS center in Basel, Switzerland. The neurologists tested whether the 200 videos were human-readable after decoding and rated the severity grade of each original and decoded video according to the Neurostatus-Expanded Disability Status Scale definitions of limb ataxia. Furthermore, the neurologists tested whether ratings were equivalent between the original and decoded videos.

Results: In total, 172 of 200 (86.0%) videos were of sufficient quality to be ratable. The intrarater agreement between the original and decoded videos was 0.317 (Cohen weighted kappa). The average difference in the ratings between the original and decoded videos was 0.26, in which the original videos were rated as more severe. The interrater agreement between the original videos was 0.459 and that between the decoded videos was 0.302. The agreement was higher when no deficits or very severe deficits were present.

Conclusions: The vast majority of videos (172/200, 86.0%) decoded by the autoencoder contained clinically relevant information and had fair intrarater agreement with the original videos. Autoencoders are a potential method for enabling the use of patient videos while preserving data privacy, especially when non–health-care professionals are involved.

(J Med Internet Res 2020;22(5):e16669) doi:10.2196/16669

KEYWORDS

autoencoder; video-rating; machine learning algorithms; deep neuronal network; Neurostatus-EDSS

Introduction

In chronic neurological diseases, especially multiple sclerosis (MS), clinical assessment of motor dysfunction is crucial to monitor the disease in patients [1]. Traditional scales used to assess MS, such as the Expanded Disability Status Scale (EDSS), are not sensitive enough to detect slight changes in motor performance [2]. Video recordings of patient performance are more accurate and increase the reliability of severity ratings [3,4]. Moreover, when these recordings are automated, quantitative disability assessments by machine learning algorithms (MLA) can be created [5]. Machine learning algorithms are potentially more sensitive in detecting small changes between images; however, they require high-resolution images because of the high dimensionality of the data [6,7]. Creation of these algorithms usually involves non–health care professionals, which is a potential challenge for maintaining data privacy. Autoencoders can address this issue. They embed visual information into a lower-dimensional latent space that preserves information needed for algorithm development but is not visually interpretable by humans. [6]. An autoencoder consists of an encoder that creates encoded videos by creating a sequence of coded frame vectors and a paired decoder that transforms the coded frame vectors back into the original video. Videos encoded in this way can be shared with non–health care professionals, while the decoder can be used to verify if the essential information from the video has been captured. However, it is unknown whether the condensed data in the coded frame vectors contain clinically relevant data. Therefore, the aim of this proof-of-concept study was to test whether coded frame vectors of autoencoders contain relevant information for analyzing videos of the motor performance of patients with MS.

Methods

Study Design and Participants

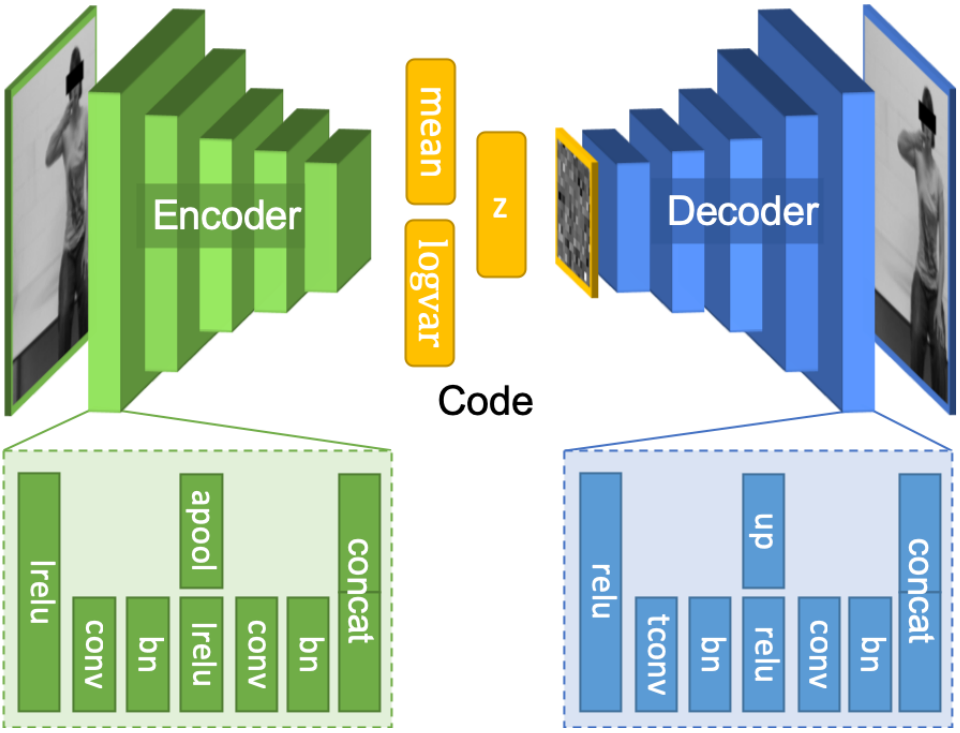
This study was a subproject of the ASSESS MS study [5] and was approved by the local ethics committees. All participants gave their written informed consent prior to inclusion. In the ASSESS MS study, 9 standardized movements were recorded on video; these movements covered overall motor function, including upper extremity function, truncal stability, and mobility. A detailed description of the movements can be found elsewhere [8]. For this study, we used recordings of the finger-to-nose test. The execution of the finger-to-nose test was standardized using a detailed protocol: Each participant was instructed to close their eyes and abduct their arms to 90° at the shoulder in full extension before touching their nose with the tip of their index finger. Both sides were tested. Original and decoded videos of 20 participants were shown to 10 neurologists at an academic MS center in Basel, Switzerland. The neurologists tested whether these 200 videos in total were human-readable after decoding and rated the severity grade of each original and decoded video according to the Neurostatus-EDSS definitions of limb ataxia [9] (subscore grade 0=no ataxia; grade 1=signs only; grade 2=tremor or clumsy movements easily seen, minor interference with function; grade 3=tremor or clumsy movements that interfere with function in all spheres; and grade 4=most functions are very difficult). The decoded videos were shown firstly, and after an interval of 2-3 weeks, the original videos were shown in the same order to minimize recall bias. The neurologists tested whether these videos were human-readable after decoding.

Autoencoder

A variational autoencoder was trained on 2230 videos comprising the 9 standardized motor performances included in the ASSESS MS study. The autoencoder was structured so that the frames of each video were encoded into a lower-dimensional space and then decoded into their original form.

Figure 1 depicts the structure of the autoencoder [10]. An encoder network was presented with a single frame from the video without further context. The frame passed through 5 encoding blocks. In each block, the input was processed in a block inspired by a densely connected convolutional network [11], wherein a skip connection was provided between the input and output layers in addition to a convolutional layer/batch normalization sequence. Each block halved the resolution of the image and doubled the feature depth. This network predicted the mean and variance of a normal distribution, which was then sampled to produce a code. The code was presented to a second network that consisted of 5 decoding blocks. Each decoding block consisted of a skip connection (which performed a simple upsampling process) and a transposed convolutional block like that used in a deep convolutional generative adversarial network [12]. Each block doubled the resolution and halved the feature depth. The network was trained using a multi-scale structural similarity–based perceptual loss function [13] with Kullback-Leibler regularization as per Kingma and Welling [10]. The input images were 256×256 RGB-D images with a code length of 256. The training hyperparameters were as follows: the learning rate was 0.001, the convolutional kernel size was 5, and the number of initial filters was 8. The model was trained for 400 epochs.

Figure 1. Structure of the variational autoencoder.



The key property of interest to us was that when a frame is in its coded form, it is computationally prohibited to decipher it without access to the decoder [6]. An autoencoder as described above reduces the dimensionality of the input data (in our case, videos) by passing the data through an "information bottleneck" [14]. The resulting coded, or latent, space sufficiently describes the data in a way that allows an accurate partial reconstruction. The shared latent embedding is optimized to represent the salient information that is similar across frames of multiple videos (in our case: the movement), whereas dissimilar aspects (eg, background aspects, details of physical features) are less well conserved. Neural networks are a machine learning approach that is inspired by biological neuronal computation; these networks have demonstrated exceptional performance in complex image-related tasks in recent years [15-17]. Given this success, in this study, we used a neural net approach called a variational autoencoder [18]. A variational autoencoder has at its center a coded vector of vastly reduced dimensionality. This is because the decoder requires millions of floating point values to be set precisely before the coded vector can be successfully decoded into an image. At the same time, the coded vector contains all the information necessary to reconstruct that frame; interestingly, due to the variational constraints during training, the frame has semantically meaningful cosine distances to other visually similar frames. This property is very useful for machine learning tasks that operate upon these coded vectors because the coded frames can be used in place of the original video frames without the possibility that a human could use it to recognize the depicted participant.

Statistics

Intrarater agreement between the ratings of the original and the decoded videos was assessed using the Cohen weighted kappa with linear weights (ie, disagreements of 1, 2, and 3 were weighted by factors of 1, 2, and 3, respectively). A Cohen kappa of 0 corresponds to chance agreement; 0-0.2, to slight agreement; 0.21-0.4, fair agreement; 0.41-0.6, to moderate agreement; 0.61-0.8, to substantial agreement; and 0.81-1, to almost perfect agreement [19]. All analyses were performed in MATLAB (MathWorks, Inc).

Results

The characteristics of the study population and the participating neurologists are summarized in Table 1.

In total, 172/200 (86.0%) videos were of sufficient quality to be ratable. The Cohen weighted kappa indicating intra-rater agreement between the original and decoded videos was 0.317. The average difference in the ratings between the original and decoded videos was 0.26, in which the original videos were rated as more severe. The inter-rater agreements of the original and decoded videos were 0.459 and 0.302, respectively. As depicted in Figure 2, agreement was higher when no deficits (grade 0) or very severe deficits (grade 4) were present. Note that most videos that were not ratable were judged so by neurologists 2 and 5.

Table 1. Characteristics of the patients and neurologists who participated in the study.

Characteristic	Value
Patient characteristics (n=20)	
Age (years), mean (95% CI)	44.4 (27-74)
Gender (female/male), n (%)	12 (63%)/7 (37%)
Disease duration (years), mean (95% CI)	13.2 (1-40)
Median EDSS[a] (range)	3.5 (0-6.5)
Type of MS[b] (RRMS[c]/SPMS[d]), n (%)	19 (95%)/1 (5%)
Neurologists (n=10)	
Gender (female/male), n (%)	5 (50%)/5 (50%)
Years of experience in neurology, mean (range)	8.8 (3 to >30)

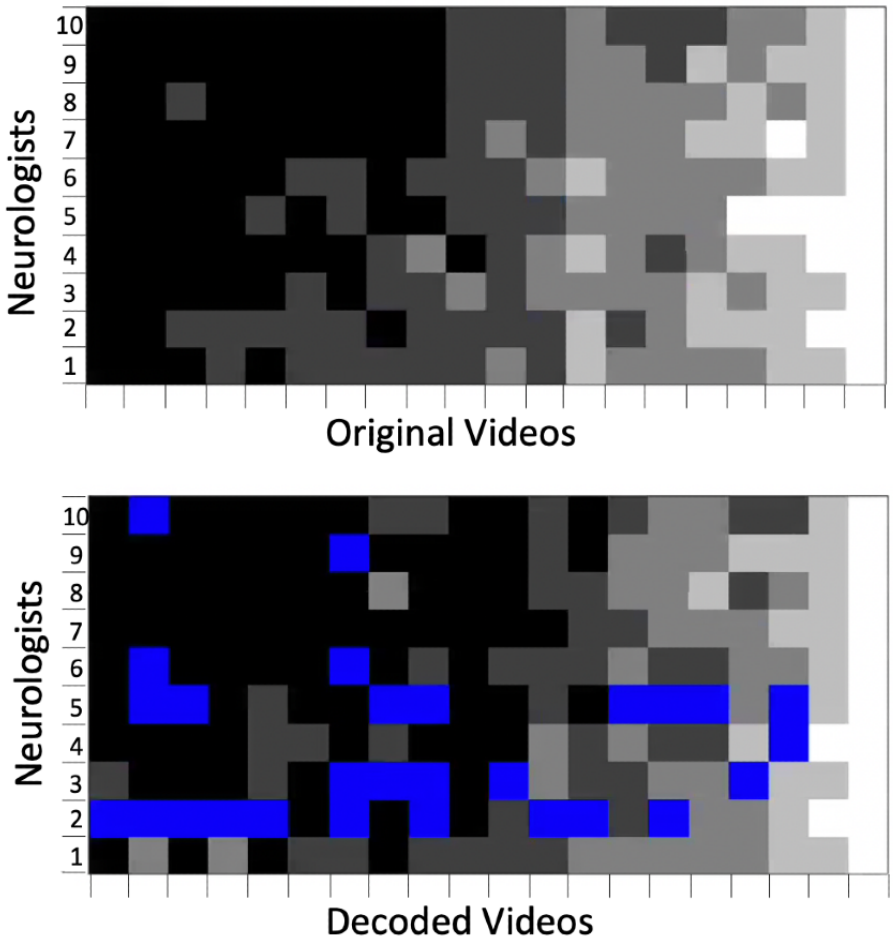
[a]EDSS: Expanded Disability Status Scale.

[b]MS: multiple sclerosis.

[c]RRMS: relapsing remitting multiple sclerosis.

[d]SPMS: secondary progressive multiple sclerosis.

Figure 2. Ratings by 10 neurologists of the original and decoded videos. The colored squares represent the different grades for limb ataxia of the finger-to-nose-test according to the Neurostatus-Expanded Disability Status Scale subscores: black=0, dark grey=1, grey=2, bright grey=3, and white=4. The blue squares represent videos that were judged as not ratable by the neurologists.



Discussion

Principal Findings

In this proof-of-concept study, 172/200 (86.0%) of the decoded videos were of sufficient quality to be ratable. We found fair intrarater agreement between the original and decoded videos. The agreement was better for minor and severe deficits in motor function.

Data security and privacy are increasingly requested by health care professionals for data capture, analysis, and storage [20]. At the same time, the use of machine learning algorithms and deep neuronal network techniques as subdomains of artificial intelligence is increasingly infiltrating all areas of health care [21,22]. The use of new technologies and electronic tools for capture and automated analysis of clinical data generally requires the involvement of non–health care professionals, which creates challenges regarding data privacy. To our knowledge, this is the first study to use an autoencoder to allow the analysis of patient videos while preserving data privacy.

Patients with MS may present with slight changes in motor performances over their disease course. Clinical assessment of these changes is notoriously difficult. Video analysis of motor performances allows automated analyses and quantification of disability by using machine learning algorithm–based analysis systems such as those used in the ASSESS MS study; however, it requires a huge data set [5]. Since the creation of machine learning algorithms usually involves non-medical collaborators, encoding of these videos is essential. The intra-rater agreement of original and decoded videos in this study was fair. It is unclear whether this is due to accordance of the video quality or the test-retest reliability of the finger-to-nose test. To our knowledge, no data are available regarding this psychometric property of the finger-to-nose test.

Limitations

A limitation of this proof-of-concept study is the class imbalance of the patient videos according to the four grades of limb ataxia for the finger-to-nose test [9,21]. Further iterations of the deep neural network are necessary to increase the intrarater reliability.

Conclusions

In this proof-of-concept study, we have shown that the vast majority (172/200, 86.0%) of videos decoded by an autoencoder contained clinically relevant information regarding upper extremity motor performance represented by the finger-to-nose test and had fair intrarater agreement. Autoencoders are a potential method for enabling the use of patient videos while preserving data privacy, especially when non–health care professionals are involved.

Acknowledgments

This study was supported by Novartis.

Conflicts of Interest

MD received travel support from Bayer AG, Biogen, Teva Pharmaceuticals, and Sanofi Genzyme and research support from University Hospital Basel. CM received travel support from Novartis Pharma AG, Sanofi Genzyme, Teva Pharmaceuticals, and Merck Serono; honoraria for lecturing and consulting from Novartis Pharma AG, Biogen-Idec, and Merck Serono; and compensation for serving on a scientific advisory board from Biogen-Idec, Roche, Merck Serono, and Sanofi Genzyme. JD is an employee of Novartis Pharma AG. AD is an employee of Novartis Pharma AG. CK has received honoraria for lectures and research support from Biogen-Idec, Novartis Pharma AG, Almirall, Bayer Schweiz AG, Teva Pharmaceuticals, Eli Lilly, Merck Serono, Sanofi Genzyme, and the Swiss Multiple Sclerosis Society. SS has received travel support from Bayer, Merck, and Novartis and has received honoraria for consulting from Bayer, Merck, Roche, and Teva. FD is an employee of Novartis Pharma AG. BU has received consultation fees from Biogen-Idec, Novartis Pharma AG, EMD Serono, Teva Pharmaceuticals, Sanofi Genzyme, and Roche. The Multiple Sclerosis Center Amsterdam has received financial support for research from Biogen-Idec, Merck Serono, Novartis Pharma AG, and Teva Pharmaceuticals. In the last 3 years, LK's institution (University Hospital Basel) received consultancy, steering committee, and advisory board fees from Actelion, Alkermes, Almirall, Bayer, Biogen, Celgene, df-mp, EXCEMED, GeNeuro SA, Genzyme, Merck, Minoryx, Mitsubishi Pharma, Novartis, Roche, Sanofi-Aventis, Santhera, Teva, and Vianex and as well as royalties for Neurostatus products. These fees were used exclusively for research support in the Department of Neurology. For educational activities of the Department, the institution received honoraria from Allergan, Almirall, Bayer, Biogen, EXCEMED, Genzyme, Merck, Novartis, Pfizer, Sanofi-Aventis, Teva, and UCB. MJ is an employee of Microsoft Research.

References

1. Cohen JA, Reingold SC, Polman CH, Wolinsky JS. Disability outcome measures in multiple sclerosis clinical trials: current status and future prospects. Lancet Neurol 2012 May;11(5):467-476. [doi: 10.1016/s1474-4422(12)70059-5] [Medline: 22516081]
2. van Munster CEP, Uitdehaag BMJ. Outcome Measures in Clinical Trials for Multiple Sclerosis. CNS Drugs 2017 Feb 9;31(3):217-236. [doi: 10.1007/s40263-017-0412-5] [Medline: 28185158]

3. Burggraaff J, Dorn J, D'Souza M, Morrison C, Kamm CP, Kontschieder P, et al. Video-Based Pairwise Comparison: Enabling the Development of Automated Rating of Motor Dysfunction in Multiple Sclerosis. Arch Phys Med Rehabil 2020 Feb;101(2):234-241. [doi: 10.1016/j.apmr.2019.07.016] [Medline: 31473205]
4. D'Souza M, Steinheimer S, Dorn J, Morrison C, Boisvert J, Kravalis K, et al. Reference videos reduce variability of motor dysfunction assessments in multiple sclerosis. Mult Scler J Exp Transl Clin 2018 Aug 09;4(3):205521731879239. [doi: 10.1177/2055217318792399] [Medline: 30116550]
5. Morrison C, D'Souza M, Huckvale K, Dorn JF, Burggraaff J, Kamm CP, et al. Usability and Acceptability of ASSESS MS: Assessment of Motor Dysfunction in Multiple Sclerosis Using Depth-Sensing Computer Vision. JMIR Hum Factors 2015 Jun 24;2(1):e11. [doi: 10.2196/humanfactors.4129] [Medline: 27025782]
6. Vieira S, Pinaya WH, Mechelli A. Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications. Neurosci Biobehav Rev 2017 Mar;74:58-75. [doi: 10.1016/j.neubiorev.2017.01.002] [Medline: 28087243]
7. Pinaya WHL, Mechelli A, Sato JR. Using deep autoencoders to identify abnormal brain structural patterns in neuropsychiatric disorders: A large - scale multi - sample study. Hum Brain Mapp 2018 Oct 11;40(3):944-954. [doi: 10.1002/hbm.24423] [Medline: 30311316]
8. van Munster CE, D'Souza M, Steinheimer S, Kamm CP, Burggraaff J, Diederich M, et al. Tasks of activities of daily living (ADL) are more valuable than the classical neurological examination to assess upper extremity function and mobility in multiple sclerosis. Mult Scler 2018 Aug 31;25(12):1673-1681. [doi: 10.1177/1352458518796690] [Medline: 30168739]
9. Kappos, L. https://www.neurostatus.net/. 2011. Neurostatus Scoring Definitions; Version 04/10.2 URL: https://www.neurostatus.net/
10. Kingma DP, Welling M. An Introduction to Variational Autoencoders. Found Trends Mach Learn 2019;12(4):307-392. [doi: 10.1561/2200000056]
11. Huang G, Liu Z, Pleiss G, Van Der Maaten L, Weinberger K. Convolutional Networks with Dense Connectivity. IEEE Trans Pattern Anal Mach Intell 2019 May 23:1-1. [doi: 10.1109/tpami.2019.2918284] [Medline: 31135351]
12. Radford A, Metz L, Chintala S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. 2016 Presented at: International Conference on Learning Representations; 2016 May 2-4; San Juan, Puerto Rico p. 1-16 URL: https://arxiv.org/abs/1511.06434
13. Wang Z, Simoncelli EP, Bovik AC. Multiscale structural similarity for image quality assessment. In: Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers. 2004 Presented at: Thirty-Seventh Asilomar Conference on Signals, Systems and Computers; 2003 Nov 9-12; Pacific Grove, CA, USA p. 1402-1402. [doi: 10.1109/acssc.2003.1292216]
14. Liou C, Huang J, Yang W. Modeling word perception using the Elman network. Neurocomputing 2008 Oct;71(16-18):3150-3157. [doi: 10.1016/j.neucom.2008.04.030]
15. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: IEEE. USA: IEEE; 2016 Dec 12 Presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA; 27-30 June 2016; as Vegas, NV, USA p. 770-778 URL: http://ieeexplore.ieee.org/document/7780459/ [doi: 10.1109/cvpr.2016.90]
16. Cao Z, Simon T, Wei S, Sheikh Y. Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields. In: IEEE. Honolulu, HI: IEEE; 2017 Presented at: IEEE Conference on Computer Vision and Pattern Recognition; 21-26 July 2017; Honolulu, HI, USA p. 7291-7299 URL: http://ieeexplore.ieee.org/document/8099626/ [doi: 10.1109/cvpr.2017.143]
17. Karras T, Aila T, Laine S, Lehtinen J. Progressive Growing of GANs for Improved Quality, Stability, and Variation. 2018 Presented at: International Conference on Learning Representations; 2018 April 30-May 3; Vancouver, BC, Canada.
18. Kingma D, Welling M. Auto-Encoding Variational Bayes. 2014 Presented at: International Conference on Learning Representations; 2014 April 14-16; Banff, Canada p. 1-14 URL: https://arxiv.org/abs/1312.6114
19. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics 1977 Mar;33(1):159-174. [doi: 10.2307/2529310] [Medline: 843571]
20. Beinke JH, Fitte C, Teuteberg F. Towards a Stakeholder-Oriented Blockchain-Based Architecture for Electronic Health Records: Design Science Research Study. J Med Internet Res 2019 Oct 7;21(10):e13585. [doi: 10.2196/13585] [Medline: 31593548]
21. Rowe M. An Introduction to Machine Learning for Clinicians. Acad Med 2019;94(10):1433-1436. [doi: 10.1097/acm.0000000000002792] [Medline: 31094727]
22. Triantafyllidis AK, Tsanas A. Applications of Machine Learning in Real-Life Digital Health Interventions: Review of the Literature. J Med Internet Res 2019 Apr 05;21(4):e12286. [doi: 10.2196/12286] [Medline: 30950797]

Abbreviations

EDSS: Expanded Disability Status Scale
MS: multiple sclerosis
RRMS: relapsing remitting multiple sclerosis
SPMS: secondary progressive multiple sclerosis

Edited by G Eysenbach; submitted 13.10.19; peer-reviewed by S Allin, A Aminbeidokhti; comments to author 06.01.20; revised version received 19.02.20; accepted 19.03.20; published 08.05.20.

Please cite as:
D'Souza M, Van Munster CEP, Dorn JF, Dorier A, Kamm CP, Steinheimer S, Dahlke F, Uitdehaag BMJ, Kappos L, Johnson M
Autoencoder as a New Method for Maintaining Data Privacy While Analyzing Videos of Patients With Motor Dysfunction: Proof-of-Concept Study
J Med Internet Res 2020;22(5):e16669
URL: https://www.jmir.org/2020/5/e16669
doi:10.2196/16669
PMID:32191621

Original Paper

Patients' Experiences and Attitudes of Using a Secure Mobile Phone App for Medical Photography: Qualitative Survey Study

Kirk D Wyatt[1], MD; Anissa Finley[2], BS; Richard Uribe[3], MBA, MHSM; Peter Pallagi[4], BFA; Brian Willaert[5], MS; Steve Ommen[6], MD; James Yiannias[6,7], MD; Thomas Hellmich[2], MD

[1]Division of Pediatric Hematology/Oncology, Department of Pediatric and Adolescent Medicine, Mayo Clinic, Rochester, MN, United States

[2]Department of Emergency Medicine, Mayo Clinic, Rochester, MN, United States

[3]Information Technology, Mayo Clinic, Scottsdale, AZ, United States

[4]Media Support Services, Mayo Clinic, Scottsdale, AZ, United States

[5]Information Technology, Mayo Clinic, Rochester, MN, United States

[6]Center for Connected Care, Mayo Clinic, Rochester, MN, United States

[7]Department of Dermatology, Mayo Clinic, Scottsdale, AZ, United States

Corresponding Author:
Kirk D Wyatt, MD
Division of Pediatric Hematology/Oncology
Department of Pediatric and Adolescent Medicine
Mayo Clinic
200 First Street Southwest
Rochester, MN, 55905
United States
Phone: 1 507 284 2695
Email: wyatt.kirk@mayo.edu

Abstract

Background: Point-of-care clinical photography using mobile devices is coming of age as a new standard of care for clinical documentation. High-quality cameras in modern smartphones facilitate faithful reproduction of clinical findings in photographs; however, clinical photographs captured on mobile devices are often taken using the native camera app on the device and transmitted using relatively insecure methods (eg, SMS text message and email) that do not preserve images as part of the electronic medical records. Native camera apps lack robust security features and direct integration with electronic health records (EHRs), which may limit patient acceptability and usefulness to clinicians. In March 2015, Mayo Clinic overcame these barriers by launching an internally developed mobile app that allows health care providers to securely capture clinical photographs and upload them to the EHR in a manner that is compliant with patient privacy and confidentiality regulations.

Objective: The study aimed to understand the perceptions, attitudes, and experiences of patients who were photographed using a mobile point-of-care clinical image capture app.

Methods: The study included a mail-out survey sent to 292 patients in Rochester, Minnesota, who were photographed using a mobile point-of-care clinical image capture app within a preceding 2-week period.

Results: The surveys were completed by 71 patients who recalled being photographed. Patients were seen in 18 different departments, with the most common departments being dermatology (19/71, 27%), vascular medicine (17/71, 24%), and family medicine (10/71, 14%). Most patients (49/62, 79%) reported that photographs were taken to simply document the appearance of a clinical finding for future reference. Only 16% (10/62) of patients said the photographs were used to obtain advice from a specialist. Furthermore, 74% (51/69) of the patients said they would recommend medical photography to others and 67% (46/69) of them thought the photos favorably affected their care. Patients were largely indifferent about the device used for photography (mobile device vs professional camera; 40/69, 58%) or the identity of the photographer (provider vs professional photographer; 52/69, 75%). In addition, 90% (64/71) of patients found reuse of photographs for one-on-one learner education to be acceptable. Acceptability for other uses declined as the size of the audience increased, with only 42% (30/71) of patients deeming reuse on social media for medical education as appropriate. Only 3% (2/71) of patients expressed privacy or confidentiality concerns. Furthermore, 52% (33/63) of patients preferred to provide consent verbally, and 21% (13/63) of them did not think a specific consent process was necessary.

Conclusions: Patient attitudes regarding medical photography using a secure EHR-integrated app were favorable. Patients perceived that photography improved their care despite the most common reason for photography being to simply document the appearance of a clinical finding for future reference. Whenever possible, health care providers should utilize secure EHR-integrated apps for point-of-care medical photography using mobile devices.

(J Med Internet Res 2020;22(5):e14412) doi:10.2196/14412

KEYWORDS

photography; mobile apps; telemedicine; electronic health records; mobile phone; digital imaging; dermatology; vascular medicine; family medicine

Introduction

Clinical Photography

Clinical photography has been standard practice in the fields of dermatology [1,2], plastic surgery [1,3] and dentistry [4] for years, and it has emerged as a useful tool for use by general practitioners [5] and emergency medicine providers [6] as well. Clinical photographs can be captured by professional medical photographers using clinic-owned equipment or by health care providers themselves using a clinic-owned camera or their personal mobile device. With personal smartphone ownership approaching nearly 3 billion, the usage of mobile devices is expected only to increase [7].

Clinical photography has many uses, and most patients seem to find it useful in the course of their care [8,9]. It can provide more vivid descriptions than a provider could detail in their written documentation, allow providers to follow progression of a disease over time, facilitate telemedicine, and allow patients or their caregivers to see clinical findings they might not otherwise be able to see themselves (eg, finding on back or intraoperative findings) [10,11].

Privacy and Confidentiality Concerns

A number of legal and ethical issues concerning patient privacy and data security arise in the course of clinical photography; for example, patients may feel uncomfortable with clinical photography that involves sensitive areas (eg, genitalia) [6]. Patient attitudes may vary when a professional-grade, clinic-owned camera is used compared with a health care provider's personal mobile phone [8,9,12,13]. Consent practices may vary [8], and the optimal method of consent (ie, implied, verbal, or written) for medical photography is unclear [2,9]. Furthermore, when patients consent to medical photography, they may assume that the photographs will only be used to provide clinical care, yet it is clear that clinical photographs may be beneficial tools for medical education [10]. The extent to which patients are comfortable with photographs being used for various educational purposes may be variable [6,8,9,12,13]. Although patients are, in general, accepting of reuse for medical care and education in small settings, they are less accepting of reuse in media that are distributed to larger audiences and the lay public (eg, social media) [6,8,9,12,13].

Legal provisions in the United States and elsewhere dictate the manner in which health care providers must keep protected health information private and secure. Although some health care providers transmit clinical photographs to colleagues using methods that lack certain privacy and security features, such as text messaging or email [2], various electronic health record (EHR) vendors have incorporated secure clinical photography modules that include patient identity management and security features within the mobile apps. Despite this, these secure mobile apps may be relatively underused. In a Canadian survey of neurosurgical residents, 45% of whom used their smartphone for clinical photography, 89% stored photos or videos of patients in the native smartphone app and only 8% used a password-protected app [9]. More concerning is that only 32% of users deleted photographs immediately after use, and 23% of them said that they do not routinely delete patient photographs on their phones [9]. At the same time, a survey of dermatologists—approximately half of whom used a smartphone for clinical photography—revealed that only 43% used a secure smartphone app that is integrated with the EHR [14]. This practice is problematic for several reasons. First, mobile devices may not be uniformly password protected to prevent unauthorized access. Second, mobile devices often automatically backup photos to a personal cloud storage service (eg, iCloud, Apple Inc; DropBox, Dropbox, Inc). These tools may not be compliant with health care data protection laws as they may lack sufficient security to prevent unauthorized viewing and there may be no reliable mechanism to ensure that patients are notified in the event of a data breech on the cloud storage service. Finally, capturing a photograph using a native app implies that the photograph will not become part of the medical record. Not only does this limit visibility by other clinicians participating in the patient's care, but it also implies that the photograph may be transmitted using insecure methods (eg, personal email, text message).

Goal

In March 2015, after legal and policy review, Mayo Clinic released an internally developed iOS-based clinical photography app named PhotoExam. The app permits any member of the health care team with clinical documentation privileges to capture photographs using their personal mobile device or a clinic-owned device and directly upload the images to the EHR. The app's features include patient identity management, Health Insurance Portability and Accountability Act (HIPAA) compliance, and confirmation that patient consent has been obtained before photography takes place. We previously reported app uptake at the 3 Mayo Clinic campuses in Minnesota, Florida, and Arizona as well as the regional Mayo Clinic Health System [11]. Considering that the existing literature largely focused on patient attitudes about medical photography performed either using a clinic-owned professional camera or a smartphone app that lacks security features and is

not integrated with the EHR, we aimed to understand how patient experience may differ with PhotoExam.

Therefore, to assess perceived patient benefit, acceptability, and privacy/security concerns, we surveyed patients who were photographed by their health care providers using the PhotoExam app.

Methods

PhotoExam App

The PhotoExam app is an internally developed iOS-based mobile app that allows health care providers to take clinical photographs of patients and incorporate them into the EHR in a manner that is secure and HIPAA compliant. Providers launch the app either from the patient's chart opened within a third-party EHR mobile app or they open the app from their home screen and then manually enter the patient's medical record. A hard-stop requires providers to verify that they have obtained patient consent, according to departmental policies, before capturing photographs. Providers then select the anatomical site(s) they will photograph and use the smartphone camera to capture up to 6 photographs per anatomical site. Photos can be immediately deleted or retaken if quality is suboptimal. After all images have been captured, the app uploads the images to the patient's medical record and provides confirmation that the images have been successfully uploaded. The photographs are then automatically deleted from the user's device after upload is complete or whenever the user closes the app—whichever comes first. Images are never made accessible outside of the PhotoExam app (eg, within the native photo gallery). Newer versions of PhotoExam allow capture of short video clips; however, video recordings were not considered for this study.

Patient Selection

We included a random sample of 300 adult patients or parents of pediatric patients (ie, aged <18 years) who were photographed using the PhotoExam app at Mayo Clinic, Rochester, Minnesota, within a preceding 2-week period and for whom a mailing address was available. Patients who refused research participation were excluded. We also excluded patients whose primary language was not English because the survey was written in English. To protect pediatric patients, we reviewed the photographs and clinical records of pediatric patients identified for potential inclusion and excluded patients who were seen for confidential visits or visits of a potentially emotionally sensitive nature. The excluded pediatric patients were substituted with another randomly selected pediatric patient (ie, aged <18 years) who had been photographed during the same 2-week period.

Survey Description

The survey included the patient's name, medical record number, and the date of the most recent clinical visit where a health care provider photographed them using the PhotoExam app. Surveys were mailed to patients at their home address. Surveys corresponding to patients younger than 18 years were addressed to the patient's parent or guardian. Patients who did not recall being photographed were asked to not answer further questions and return the survey. Questions generally related to the reason photos were taken, effect on timeliness of care, comfort level with medical photography, and the manner in which consent to photography was obtained. Patients were also allowed to share general comments or concerns about the practice of medical photography using mobile devices in free-text form (Multimedia Appendix 1). Patients who did not respond to the initial survey were sent a follow-up survey.

Statistical Analysis

Responses are reported in counts and percentages with continuous variables summarized as means or medians and standard deviations or ranges, as appropriate. In cases where patients did not respond to individual questions, percentages represent the number of patients who responded to an individual question.

Human Subjects Protection

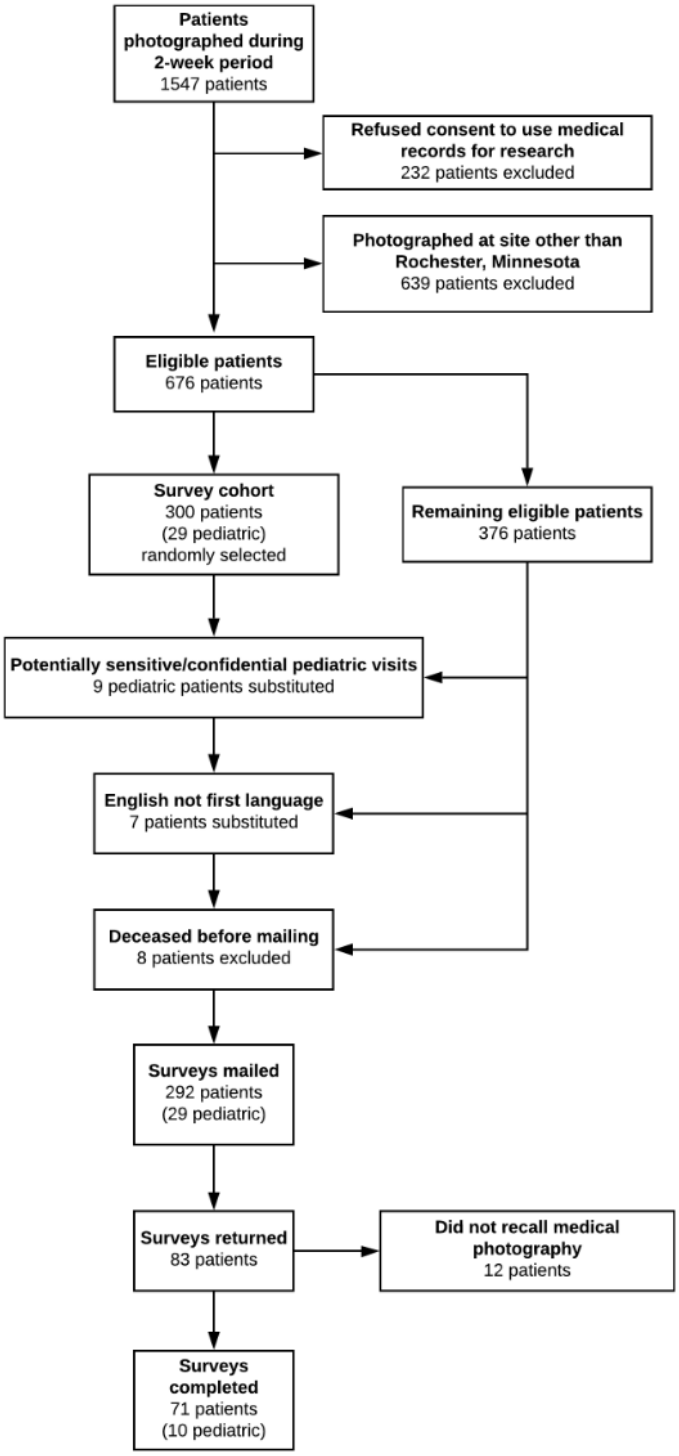
The study procedures were approved by the Mayo Clinic Institutional Review Board.

Results

Patient Selection

Patients (n=1547) who were photographed with the app within a 2-week period were identified. After ineligible patients were excluded, surveys were printed and mailed to 292 patients, 29 of whom were pediatric patients (Figure 1). Surveys were mailed 51 days after the end of the 2-week period when photos were taken. A total of 64 completed surveys were returned after the initial mailing. The second mailing was sent to nonrespondents 1 month after the initial mailing (81 days after the end of the 2-week period where photos were taken). A total of 83 total surveys were returned. In addition, 14% (12/83 of respondents willing to complete the survey) adult patients did not recall a health care provider using a mobile device to take a photograph during a visit and therefore did not complete the remainder of the survey. The remainder of respondents (n=71; 10 pediatric) recalled medical photography taking place and completed the remainder of the survey.

Figure 1. Patient selection.



Population and Demographics

The patient demographics are shown in Table 1. Patients were seen in 18 different departments, with the most common departments being dermatology (19/71, 27%), vascular medicine (17/71, 24%), and family medicine (10/71, 14%).

Table 1. Patient demographics (N=83 surveys returned).

Characteristics	Values
Age (years), median (SD)	63 (23)
Younger than 18 years of age, n (%)	10 (12)
Marital status (N=83 surveys returned), n (%)	
Single	19 (23)
Married	59 (71)
Divorced	4 (5)
Widowed	1 (1)
Race (N=83 surveys returned), n (%)	
White	82 (99)
Other	1 (1)
Number of photos taken of patient using PhotoExam during an encounter, median (range)	2 (1-9)
Number of photos ever taken of patient using PhotoExam, all encounters included, median (range)	5 (1-164)
Days between clinical visit and survey completion (days), median (range)	68 (58-224)
Eligible surveys (N=292 surveys mailed), n (%)	
Survey returned	83 (28.4)
No response after 3 mailings	177 (60.6)
Refused	24 (8.2)
Deceased (survey not sent)	8 (2.7)
Survey returned without Health Insurance Portability and Accountability Act authorization	3 (1.0)
Deceased (notified after survey sent)	2 (0.6)
Invalid mailing address	1 (0.3)
Physically/mentally unable to complete	1 (0.3)
Returned blank survey	1 (0.3)

Rationale for Photography

Patients reported that the majority of photos were taken to document the appearance of the area for future reference (49/71, 69%). Furthermore, 14% (10/71) photos were taken to send to a specialist for review. Three percent (2/71) reported that photographs were taken for educational purposes, and the remainder were unsure why photographs were taken.

Effect of Photography on Time to Diagnosis and Treatment When Specialist was Consulted

Two of the patients whose providers photographed them to obtain a specialist's assistance with making the diagnosis said the photography had no effect on the time to diagnosis and treatment, but the remaining 3 indicated that it either slightly or significantly expedited time to diagnosis and treatment.

Consent for Photography

Overall, 91% (60/66) of patients who were surveyed recalled being asked permission to take photographs, with three-fourth (45/60) of these patients providing verbal permission and the rest (15/60) providing written permission.

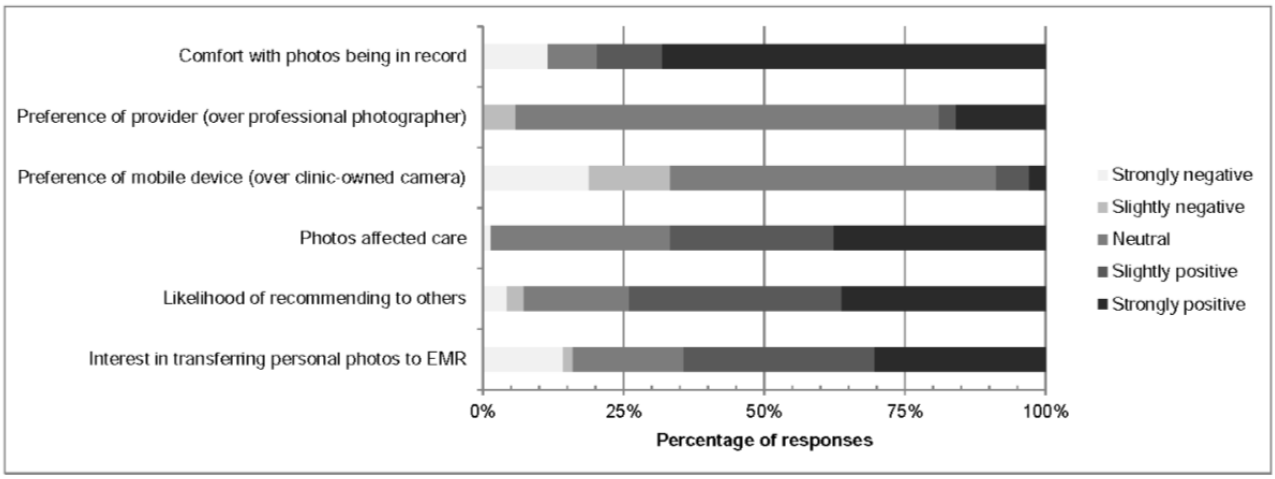
Of the patients who provided written consent, 53% (8/15) indicated that they only read a part of the consent form and 1 patient (1/15, 7%) did not read any part of the form, with the remaining 33% (5/15) patients indicating that they read the entire consent form word for word.

When patients were asked their preference for providing consent, 21% (13/63) did not think specific permission was necessary, 52% (n= 33/63) preferred verbal consent, and 27% (17/63) preferred to provide written consent, with 59% (10/17) of these patients requesting to provide written consent by signing the screen on the device used to take photos (a feature that was not available).

Attitudes Regarding Photography Practice and Process

Patients were asked a variety of questions regarding their attitudes about photography, including their comfort with photos being present in the EHR, preference of photographer, and preference of photography equipment. Patients were also asked whether photography affected their care and whether they would recommend photography to family or friends in a similar clinical scenario. Responses are shown in Figure 2, where Likert scale responses are color coded. Overall, 74% (51/69) of patients said they were likely to recommend mobile point-of-care clinical photography to others, and 67% (46/69) of patients thought the photos favorably affected their care.

Figure 2. Attitudes regarding photography practice and process. EMR: electronic medical record.



Furthermore, 80% (56/70) of patients indicated that they owned a phone with a camera, and the majority of them expressed an interest in transferring personally taken photographs into the EHR (Figure 2). Three of the patients (3/55, 5%) who owned a smartphone experienced a provider taking a photograph of a phone screen to transfer a photograph from a personal device to the EHR.

Privacy and Data Security Concerns

In total, 2 patients (2/71, 3%) indicated that they had concerns regarding privacy or confidentiality related to the photos. When asked for further explanation, one of these patients inquired whether the photos remain on the provider's phone and another explained that they believed the photos were used professionally but that they did not know the provider who took the photos. Therefore, it appeared that these concerns did not represent a belief that a privacy breech had actually occurred.

In addition, 26% (18/70) of patients recalled that security features of the app were explained by the health care provider. Furthermore, 16% (11/70) indicated that security features were not explained, and the majority (41/70, 59%) of them did not remember or were unsure if security features were explained to them.

Reuse of Photographs for Educational Purposes

Patients were asked about their comfort level with the photos that were taken being shared in various settings, provided that any personally identifying information was removed. In general, patients were supportive of use for medical education; however, comfort declined with progressively increasing audience sizes (Table 2).

Table 2. Patient comfort with reuse of photographs for educational purposes (N=71)

Reuse of photographs	Value, n (%)
One-on-one learner education	64 (90)
Large group (eg, classroom) medical learner education	57 (80)
Presentation at a national medical professional meeting	52 (73)
Publication in a medical journal or textbook	48 (68)
Publication on social media for medical education	30 (42)

Free-Text Responses

Overall, 41% (29/71) of patients provided additional written comments regarding their experiences with being photographed using a mobile device. The comments included the following:

- Requests to have a personal copy of the photographs sent.
- Indication that answers regarding comfort with the photos would have been different had the anatomical site photographed been a more sensitive area.
- Indication that the ability to have medical photographs taken prevented the need for an additional office visit.
- General questions regarding the security of the app and desire for a patient education handout regarding the privacy and security of the app.
- Indication that the ability to review images allowed patients to track the progress of their medical condition.

Discussion

Principal Findings

To the best of our knowledge, this is the first report to focus on the attitudes and experiences of patients who were photographed using a secure EHR-integrated smartphone app designed specifically for medical photography. Patients were largely satisfied with clinical photography using the app, with 74% (51/69) of patients indicating they were likely to recommend a friend or family member to give permission for medical photography using a mobile device in a similar clinical situation.

In addition, 67% (46/69) of patients perceived that the app's use improved the care they received even though most patients indicated that the photograph was taken only to document the appearance of a clinical finding that could be referenced at a future visit. This suggests that patients see benefit in the documentation of findings in a manner that may facilitate future clinical care. Indeed, in free-text responses, 2 patients shared that it helped them to track the progress of their medical condition. Although a minority of photographs were sent to a specialist, some patients shared that sending photographs to a specialist prevented them from needing an additional office visit—presumably because of the use of telemedicine.

Responses demonstrated that some patients had a desire to incorporate personally taken photographs into their medical record and revealed evidence that some health care providers had resorted to using PhotoExam to take a photograph of a photograph displayed on a patient's smartphone screen as a workaround to incorporate these photographs. EHR vendors should consider how to better integrate photograph upload functionality into their patient-facing products so that patient-taken photographs may complement those taken by providers. After the survey was administered, Mayo Clinic implemented a new commercially vended EHR that included the ability of patients to upload clinical photographs to providers with whom they have an established relationship. In the future, we will explore whether patients could be permitted to upload photographs to providers who are not already members of their care team (eg, emergency medicine provider).

The consent processes were variable. Although many patients feel that verbal consent is adequate, verbal consent processes may be problematic in that they may not elicit specific permission for all potential reuses (ie, medical education). Furthermore, even though written consent forms may ask for permission to use photographs for medical education, they may not explicitly detail various medical education uses for which patients may consent (eg, internal use one-on-one with learners vs widely distributed patient education pamphlet) and may simply include a boilerplate legal *blanket statement* seeking permission for all possible uses. We are currently evaluating the use of a 3-tiered consent form with clear language, allowing patients to consent for use of photographs for (1) clinical care only; (2) clinical care and internal education; or (3) clinical care, internal education, and external education.

Comparison With Prior Work

In our survey, most patients who recalled the consent process indicated that they provided verbal consent, and the majority of patients felt that verbal consent was adequate. In contrast, a French survey involving 158 adult patients and parents of 114 pediatric patients photographed in a dermatology clinic observed that 80% of adult patients and 89% of parents thought written consent was necessary for medical photography [8]. Another study surveyed a convenience sample of 398 patients seen in dermatology clinics in New York City and observed that 78.4% (312/398) of respondents preferred to provide written consent compared with 14.1% (56/398) who preferred verbal consent. A similar study conducted in Chicago reported 58.7% (172/293) preferring written consent over verbal consent (113/293, 38.6%) [13]. Potential explanations for the differences observed between our survey and those reported in the literature include temporal changes in patient attitudes between the survey time points (with patients gaining acceptability over time); patients' understanding of security features in the PhotoExam app; and sociodemographic differences between study populations, including the enrichment of our community with health care workers.

Patients were overwhelmingly supportive of reuse of their clinical photographs for educational purposes. Acceptability of different uses decreased as the size of the audience increased. These observations were consistent with other studies that generally reported high rates (ie, >80%) of acceptability of use of photographs for medical teaching and less comfort with widespread distribution to the lay public [6,8,9,12,13]. Despite these high levels of acceptability of image reuse, consent for reuse should always be sought (either at the time of photography or at the time reuse is desired).

We also observed, consistent with other studies, that comfort with photographs depends on the anatomical area photographed, with patients expressing discomfort with photography of sensitive anatomic areas [8,9]. A total of 3% (2/71) of actual photos in our sample were of the breast, buttock, or genitalia, and free-text comments suggested that other patients would have expressed more discomfort if they had been photographed in a more anatomically sensitive area.

More than half of patients had no preference regarding whether a provider's personal mobile device was used for clinical photography compared with a clinic-owned camera. This is in contradiction to multiple previous studies that reported poor acceptability of providers' personal mobile devices for clinical photography [8,9,12,13]. As with the differences observed regarding attitudes about the consent process in other studies, it is unclear if changes in patients' perceptions over time, the security features of the PhotoExam app, or sociodemographic differences account for the inconsistency between studies.

Strengths and Limitations

This study has several strengths. First, we included a random sample of pediatric and adult patients seen in a variety of departments who were photographed using a mobile app designed specifically for the purpose of clinical medical photography. Surveys were mailed shortly after the clinical encounter where photography occurred to maximize recall. In addition, we surveyed patients across multiple domains, including the perceived benefit, attitudes regarding use of photographs for education, preferred method of providing consent, and preferred camera type. We were also able to incorporate additional information from the medical record, including patient demographics, anatomical site(s) photographed, and the department of the specialist who photographed the patient.

Conversely, there are several limitations to this study. Although initial surveys were sent within 10 weeks of the occurrence of medical photography and most completed surveys were filled out promptly, we were surprised that 14% (12/82) of those who returned the survey did not recall that they were photographed

by their health care provider. This significantly limited our ability to gather information about these patients' experience with medical photography. It is unclear whether patients' failure to recall the photography is evidence that medical photography was viewed as an insignificant and unimportant part of their visit or whether it was perceived as standard of care and therefore not memorable. Although our response rate was low, at 28% (83/292), this was similar to response rates of similar internal patient surveys conducted at Mayo Clinic. The low response rate led to a small sample size, which limits the generalizability of the results and limits our ability to make meaningful conclusions regarding questions that only a subset of patients were eligible to respond to based on other responses.

In addition, responses were only gathered from patients, and clinical outcome measurement was limited to a patient's perception of how medical photography affected their care. Use of a formally validated survey would improve the internal and external validity of the study. Health care providers who use clinical photographs may have other insights into how clinical photography affects clinical management. To address provider experience, we plan to separately report a survey of health care providers who have used PhotoExam.

It is also important to note that our sample is biased in that we only surveyed patients who were photographed using the PhotoExam app. By nature of permitting a health care provider to photograph them with the app, patients may have preconceived notions of how the app may be beneficial or may have different attitudes regarding medical photography when compared with patients who refused medical photography with the PhotoExam app. In the authors' personal experience, refusal of consent to photograph is rare and is generally in the context of a request to photograph an anatomically sensitive area. As noted above, the patient population seen in Rochester, Minnesota, may not be representative of the broader United States. Although our population includes both local patients and those who travel from elsewhere to receive care, our local population is enriched with health care workers, whose attitudes may differ from those who do not work in health care; for example, 82% (68/83) of patients in our study were from Minnesota, with 47% (39/83) of included patients residing in Rochester, Minnesota, and an additional 7% (6/83) living outside of Rochester, Minnesota, but within Olmsted County. In addition, our patient population was 99% (82/83) white. A more racially diverse cohort may have resulted if we had provided translated surveys for non–English-speaking patients. Owing to this limitation, our current results may have failed to reflect important cultural differences in attitudes that may vary among patients.

We were encouraged that, in general, patients felt that medical photography using the app favorably affected their care, and only on a rare occasion did patients perceive it to be detrimental to their care. Concerns about privacy were rare, and patients were generally comfortable with clinical photography conducted using their provider's mobile device. These findings suggest that PhotoExam was implemented in a way that facilitates patient care and is sensitive to patients' privacy and confidentiality.

Conclusions

In summary, the PhotoExam app was well received by patients, addressed privacy and confidentiality issues, and was perceived to favorably affect patient care. We discovered no major patient-perceived barriers to implementing point-of-care clinical photography. In this regard, point-of-care medical photography is an up-and-coming best practice that we predict will become a new standard of care because of the obvious clinical benefit and ease with which photographs can be captured and entered into the medical record [15]. Future priorities include explicitly clarifying permissible educational reuses of photographs using uniform consent processes.

Acknowledgments

The authors thank the Mayo Clinic Center for Connected Care for funding to conduct the surveys. SO and JY are members of the Center for Connected Care leadership. JY provided minor suggestions regarding wording of survey questions; however, no member of Mayo Clinic Center for Connected Care otherwise played a role in the content of the survey questions. SO and JY were invited to participate as coauthors and both reviewed and approved the manuscript for publication. However, the funding source had no role in the decision to write this manuscript or submit it for publication. Furthermore, the authors thank the patients who participated in this survey and the Mayo Clinic Survey Research Center for assistance mailing surveys and compiling results.

Conflicts of Interest

All coauthors are current or former employees at Mayo Clinic, which retains intellectual property of the PhotoExam app. The app is only used internally and is not currently being licensed. Funding to distribute surveys and enter survey data was provided by Mayo Clinic Center for Connected Care, for which SO and JY are leadership members.

Multimedia Appendix 1
Patient survey.
[PDF File (Adobe PDF File), 1278 KB - jmir_v22i5e14412_app1.pdf]

References

1. Aveta A, Filoni A, Persichetti P. Digital photography in plastic surgery: the importance of standardization in the era of medicolegal issues. Plast Reconstr Surg 2012 Sep;130(3):490e-1e; author reply 491e. [doi: 10.1097/PRS.0b013e31825dc515] [Medline: 22929285]
2. Kunde L, McMeniman E, Parker M. Clinical photography in dermatology: ethical and medico-legal considerations in the age of digital and smartphone technology. Australas J Dermatol 2013 Aug;54(3):192-197. [doi: 10.1111/ajd.12063] [Medline: 23713892]
3. Yuce Y, Erkal KH, Klavuz O. Our experiences in medical photography at the burn center. Ortadogu Med J 2019;11(2):161-166 [FREE Full text] [doi: 10.21601/ortadogutipdergisi.369473]
4. Sandler J, Gutierrez RJ, Murray A. Clinical photographs: the gold standard, an update. Prog Orthod 2012 Nov;13(3):296-303. [doi: 10.1016/j.pio.2011.12.002] [Medline: 23260541]
5. Morse GA, Haque MS, Sharland MR, Burke FJ. The use of clinical photography by UK general dental practitioners. Br Dent J 2010 Jan 9;208(1):E1; discussion 14-E1; discussion 15. [doi: 10.1038/sj.bdj.2010.2] [Medline: 20057431]
6. Cheung A, Al-Ausi M, Hathorn I, Hyam J, Jaye P. Patients' attitudes toward medical photography in the emergency department. Emerg Med J 2005 Aug;22(8):609 [FREE Full text] [doi: 10.1136/emj.2004.019893] [Medline: 16046785]
7. Statista. 2019. Mobile Phone User Penetration as Percentage of the Population Worldwide From 2013 to 2019 URL: https://www.statista.com/statistics/470018/mobile-phone-user-penetration-worldwide [accessed 2019-04-16] [WebCite Cache ID 77gG7GJ04]
8. Hacard F, Maruani A, Delaplace M, Caille A, Machet L, Lorette G, et al. Patients' acceptance of medical photography in a French adult and paediatric dermatology department: a questionnaire survey. Br J Dermatol 2013 Aug;169(2):298-305. [doi: 10.1111/bjd.12345] [Medline: 23551168]
9. Leger MC, Wu T, Haimovic A, Kaplan R, Sanchez M, Cohen D, et al. Patient perspectives on medical photography in dermatology. Dermatol Surg 2014 Sep;40(9):1028-1037. [doi: 10.1097/01.DSS.0000452632.22081.79] [Medline: 25099296]
10. Nik-Hussin NM, Saleem Y, Sivayoham E, Rothera MP. A survey of parent's attitudes towards viewing intraoperative photographs used as an educational tool. Int J Pediatr Otorhinolaryngol 2009 Apr;73(4):585-588. [doi: 10.1016/j.ijporl.2008.12.029] [Medline: 19250689]
11. Wyatt KD, Willaert BN, Pallagi PJ, Uribe RA, Yiannias JA, Hellmich TR. PhotoExam: adoption of an iOS-based clinical image capture application at Mayo Clinic. Int J Dermatol 2017 Dec;56(12):1359-1365. [doi: 10.1111/ijd.13648] [Medline: 28497467]
12. Lau CK, Schumacher HH, Irwin MS. Patients' perception of medical photography. J Plast Reconstr Aesthet Surg 2010 Jun;63(6):e507-e511. [doi: 10.1016/j.bjps.2009.11.005] [Medline: 19959408]
13. Hsieh C, Yun D, Bhatia AC, Hsu JT, de Luzuriaga AM. Patient perception on the usage of smartphones for medical photography and for reference in dermatology. Dermatol Surg 2015 Jan;41(1):149-154. [doi: 10.1097/DSS.0000000000000213] [Medline: 25533160]
14. Milam EC, Leger MC. Use of medical photography among dermatologists: a nationwide online survey study. J Eur Acad Dermatol Venereol 2018 Oct;32(10):1804-1809. [doi: 10.1111/jdv.14839] [Medline: 29405432]
15. Harting MT, DeWees JM, Vela KM, Khirallah RT. Medical photography: current technology, evolving issues and legal perspectives. Int J Clin Pract 2015 Apr;69(4):401-409. [doi: 10.1111/ijcp.12627] [Medline: 25708063]

Abbreviations

EHR: electronic health record
HIPAA: Health Insurance Portability and Accountability Act

Edited by G Eysenbach; submitted 16.04.19; peer-reviewed by C Reis, A Jodheea-Jutton, Z Mungloo-Dilmohamud; comments to author 14.12.19; revised version received 25.01.20; accepted 22.02.20; published 12.05.20.

Please cite as:
Wyatt KD, Finley A, Uribe R, Pallagi P, Willaert B, Ommen S, Yiannias J, Hellmich T
Patients' Experiences and Attitudes of Using a Secure Mobile Phone App for Medical Photography: Qualitative Survey Study
J Med Internet Res 2020;22(5):e14412
URL: http://www.jmir.org/2020/5/e14412/
doi:10.2196/14412
PMID:32396127

Original Paper

Real-Time Monitoring of Blood Pressure Using Digitalized Pulse Arrival Time Calculation Technology for Prompt Detection of Sudden Hypertensive Episodes During Laryngeal Microsurgery: Retrospective Observational Study

Yong-Seok Park[1], MD; Sung-Hoon Kim[1], MD, PhD; Yoon Se Lee[2], MD, PhD; Seung-Ho Choi[2], MD, PhD; Seung-Woo Ku[1], MD, PhD; Gyu-Sam Hwang[1], MD, PhD

[1]Department of Anesthesiology and Pain Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea
[2]Department of Otolaryngology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea

Corresponding Author:
Sung-Hoon Kim, MD, PhD
Department of Anesthesiology and Pain Medicine
Asan Medical Center
University of Ulsan College of Medicine
88, Olympic-Ro 43-Gil, Songpa-Gu
Seoul, 05505
Republic of Korea
Phone: 82 2 3010 0617
Fax: 82 2 3010 6790
Email: shkimans@amc.seoul.kr

Abstract

Background: Laryngeal microsurgery (LMS) is often accompanied by a sudden increase in blood pressure (BP) during surgery because of stimulation around the larynx. This sudden change in the hemodynamic status is not immediately reflected in a casual cuff-type measurement that takes intermittent readings every 3 to 5 min.

Objective: This study aimed to investigate the potential of pulse arrival time (PAT) as a marker for a BP surge, which usually occurs in patients undergoing LMS.

Methods: Intermittent measurements of BP and electrocardiogram (ECG) and photoplethysmogram (PPG) signals were recorded during LMS. PAT was defined as the interval between the R-peak on the ECG and the maximum slope on the PPG. Mean PAT values before and after BP increase were compared. PPG-related parameters and the correlations between changes in these variables were calculated.

Results: BP surged because of laryngoscopic manipulation (mean systolic BP [SBP] from 115.3, SD 21.4 mmHg, to 159.9, SD 25.2 mmHg; $P<.001$), whereas PAT decreased significantly (from mean 460.6, SD 51.9 ms, to 405.8, SD 50.1 ms; $P<.001$) in most of the cases. The change in SBP showed a significant correlation with the inverse of the PAT ($r=0.582$; $P<.001$). Receiver-operating characteristic curve analysis indicated that an increase of 11.5% in the inverse of the PAT could detect a 40% increase in SBP, and the area under the curve was 0.814.

Conclusions: During LMS, where invasive arterial catheterization is not always possible, PAT shows good correlation with SBP and may, therefore, have the potential to identify abrupt BP surges during laryngoscopic manipulations in a noninvasive manner.

(J Med Internet Res 2020;22(5):e13156) doi:10.2196/13156

KEYWORDS

larynx; blood pressure; photoplethysmography; pulse

Introduction

Background

Laryngeal microsurgery (LMS) is often accompanied by a sudden increase in blood pressure (BP) during surgery because of stimulation around the larynx [1,2]. Laryngoscopic manipulations cause sympathetic nervous system stimulation and increase in the level of plasma catecholamines, such as epinephrine and norepinephrine [3]. Acute hypertension resulting from this process can cause complications such as ischemic heart disease, heart failure, stroke, and life-threatening arterial bleeding even in patients without predisposing factors [4,5]. Moreover, myocardial ischemia or arrhythmia can occur during the perioperative period because of the stimulation of the deep receptors of both the larynx and cardioinhibitory fibers of the vagus nerve [6]. There have been previous studies regarding the appropriate methods of anesthesia for maintaining a stable hemodynamic status in patients during LMS [7,8]. Despite a sudden change in the hemodynamic status, an abrupt BP rise is not immediately reflected in a casual cuff-type measurement that takes intermittent readings every 3 to 5 min.

Pulse arrival time (PAT), measured as the interval from the R-peak on an electrocardiogram (ECG) to the peripheral arrival of the pulse wave, has been used as an estimate for pulse transit time and may be a noninvasive marker for BP [9-14]. Recent advancements in technology enabled adjustment for confounding factors (eg, heart rate, arterial stiffness, and pre-ejection time) in estimating BP from PAT [9,15,16]. Furthermore, a cuff-less approach for 24-hour BP monitoring became possible using wearable devices [13]. Clinical application and commercialization of these wearable BP-monitoring devices are underway [17,18].

Objectives

We previously reported that beat-to-beat changes in PAT can effectively detect decreases in systolic BP (SBP) during anesthesia induction in hypertensive patients undergoing renal transplantation [11]. However, most previous studies have estimated the state of hypotensive events, but no report to date has focused on PAT monitoring to detect an abrupt BP surge. In this study, we investigated whether noninvasive PAT could be used to monitor rapid BP increases during LMS in real time.

Methods

Study Population and Anesthesia Protocol

We retrospectively analyzed the electronic records of 30 patients with vocal cord polyps or edema who were scheduled for an elective LMS under general anesthesia. Patients with cardiac arrhythmias or incomplete data were excluded.

All patients were prepared for general anesthesia according to our institutional protocol. Premedication was not given. Cardiovascular medication, including antihypertensive medication, was administered until the day of surgery, except for angiotensin II receptor blocker or angiotensin-converting enzyme inhibitor. Routine monitoring included noninvasive intermittent BP (NIBP), ECG, pulse oximetry, and end-tidal concentration of CO_2 using a multiparameter monitor (Philips IntelliVue MP70; Philips). All data were recorded simultaneously throughout the procedure. Volatile induction and maintenance of anesthesia were performed. Anesthesia was induced with 6% to 8% sevoflurane and rocuronium (0.5-1 mg/kg) and maintained with 2% sevoflurane and 50% nitrous oxide in oxygen.

Data Acquisition and Signal Processing

ECG and photoplethysmogram (PPG) waveform data were recorded with data acquisition software (Vital Recorder) [19] at a sampling rate of 300 Hz, and NIBP was measured and recorded every 3 min. The signal data were transferred via serial port from the MP70 monitor to a computer running the Vital Recorder and written to the hard drive. PAT was defined as the interval between the R-peak on the ECG and the point at which the maximal rising slope appears on the PPG [11]. The PAT of each beat was calculated using the filter function, which was built using Python code in the Vital Recorder software (Multimedia Appendix 1). When the filter function is executed, the PAT at every heartbeat is calculated from the recorded ECG and PPG waveforms by the algorithm, and the PAT values at each time point are added as a new time series variable. The PPG waveform data recorded with the data acquisition software were converted to European Data Format [20] and analyzed with signal processing software (LabChart 8; AD Instruments) to acquire features of the PPG waveform, such as height, $width_{50}$ (width at 50% height of each PPG wave peak), maximum slope, minimum slope, and area of each wave peaks [21]. These features were extracted by the peak analysis function of the software. The 1-min averages of each value just before and after laryngoscopic manipulation were calculated by the function and recorded for further analysis. All the acquired parameters during the laryngoscopic manipulation were compared. The 1-min averages of PAT just before laryngoscope insertion and at the lowest point of PAT after stimulation were used as the PAT parameters for each time point. PPG and PAT parameters were obtained simultaneously.

Statistical Analysis

All study data are presented as mean (SD), n (%), or median (IQR). A Shapiro-Wilk test was used as a test of normality. A paired *t* test or Wilcoxon signed-rank test was used to compare parameters before and after LMS. The correlation between the changes in BP and PAT and between the changes in BP and PPG-related variables at all time points were evaluated with a Pearson correlation coefficient or Spearman correlation coefficient, as appropriate. Receiver operating characteristic (ROC) curve analyses were performed to evaluate the ability of PAT and PPG-related variables to detect a 40% change in the SBP. R version 3.4.2 software (R Foundation for Statistical Computing) was used for all statistical analyses.

Ethics Approval and Consent to Participate

This study was approved by the institutional review board at Asan Medical Center (IRB No, 2017-2268). The requirement for written informed consent was waived because of the retrospective nature of the study and the anonymity of the biosignals used in this study.

Availability of Data and Materials

All data generated or analyzed during this study are included in this published paper and Multimedia Appendices 1 and 2.

Results

Changes in Hemodynamic Variables and Photoplethysmogram Variables

After data processing, 60 sets of BP, ECG, and PPG values were analyzed. The demographic data of study patients are presented in Table 1. A representative plot of SBP and PAT of a randomly selected patient is shown in Figure 1. As indicated in Figure 1, systolic NIBP showed a tendency to follow the changing trends in the inverse of the PAT at intervals of 1 to 2 min, as the cuff-type NIBP device measures arterial pressure intermittently. Laryngoscopic manipulation during LMS caused a mean 41.3% increase in SBP and a mean 13.4% decrease in PAT. The inverse of the PAT increased as SBP increased in all patients except one (Figure 2). Heart rate increased by an average of 31.3% from baseline, and morphological parameters of PPG (eg, maximum slope, minimum slope, width, and area) also changed significantly after laryngoscope insertion (Table 2).

Table 1. Demographic and clinical characteristics of the 30 study patients.

Patient characteristics and prior comorbidities	Values
Sex, male, n (%)	18 (60)
Age (years), mean (SD)	58 (14)
Height (cm), mean (SD)	161.2 (7.6)
Weight (kg), mean (SD)	65.3 (12.2)
BMI (kg/m^2), mean (SD)	24.2 (2.9)
Hypertension, n (%)	12 (40)
Diabetes mellitus, n (%)	4 (13)
Calcium channel blocker, n (%)	9 (30)
Beta blocker, n (%)	2 (6)
Angiotensin II receptor blocker, n (%)	3 (10)

Figure 1. Tracing showing the relationship between systolic noninvasive intermittent blood pressure and pulse arrival time during laryngeal microscopic surgery. Beat-to-beat changes in pulse arrival time instantaneously reflected marked perturbations of systolic noninvasive intermittent blood pressure during endotracheal intubation and the beginning of the laryngeal microsurgery procedure. The insets show the electrocardiography (upper) and the photoplethysmographic waveform (bottom) results before and after the insertion of the laryngeal microscope. The thick blue line is the smoothed line of the PAT using locally estimated scatterplot smoothing.

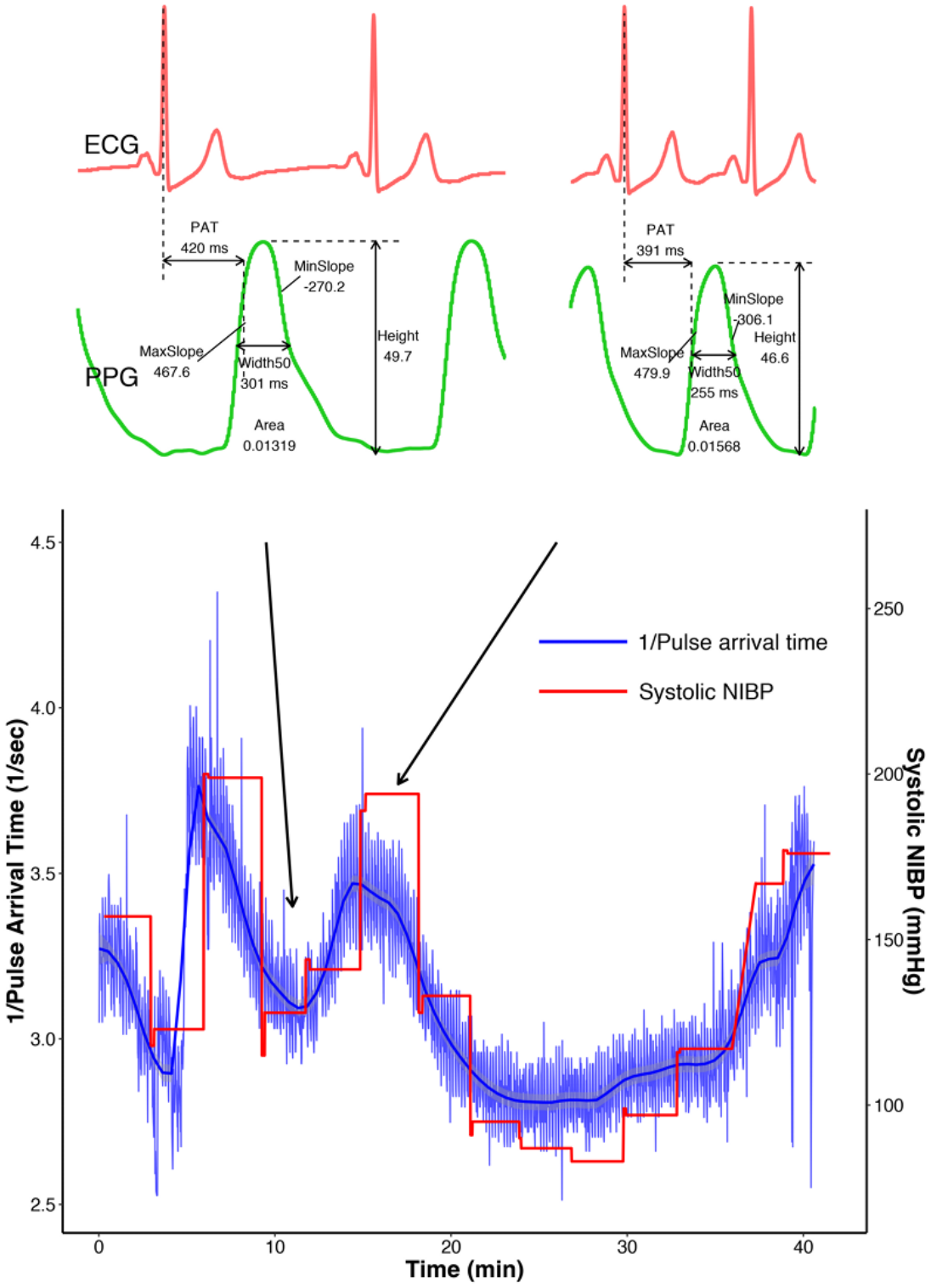


Figure 2. Changes in systolic blood pressure and pulse arrival time before and after surgical stimulation because of laryngeal microsurgery. Red dots represent changes before stimulation, whereas blue dots represent changes after stimulation.

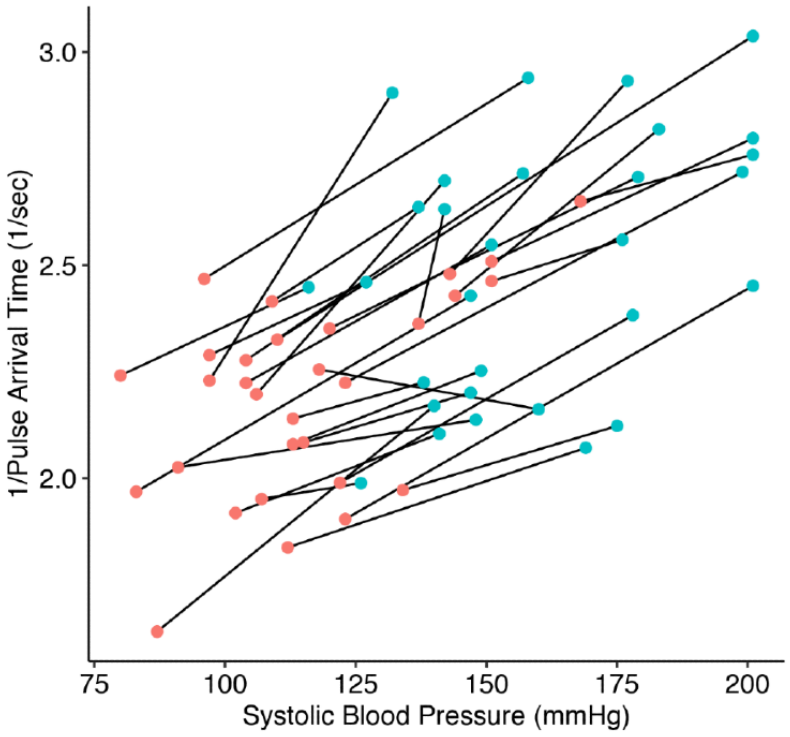


Table 2. Changes in hemodynamic, photoplethysmographic, and pulse arrival time variables during blood pressure surge (N=30).

Variables	Before	After	*P* value
Systolic blood pressure (mm Hg), mean (SD)	115.3 (21.4)	159.9 (25.2)	<.001
Diastolic blood pressure (mm Hg), mean (SD)	69.0 (14.2)	97.6 (15.6)	<.001
Mean arterial blood pressure (mm Hg), mean (SD)	79.8 (14.7)	113.2 (17.0)	<.001
Heart rate (bpm), mean (SD)	78.1 (11.7)	101.3 (12.3)	<.001
Pulse arrival time (ms), mean (SD)	460.6 (51.9)	405.8 (50.1)	<.001
PPG[a] amplitude (AU[b]), median (IQR)	45.6 (44.6-46.8)	46.3 (45.0-47.0)	.49
PPG maximum slope (AU/s), mean (SD)	485.0 (81.9)	564.3 (81.1)	<.001
PPG minimum slope (AU/s), mean (SD)	−247.2 (49.8)	−337.9 (50.3)	<.001
PPG $width_{50}$ (ms), median (IQR)	256.8 (233.3-278.4)	221.9 (209.4-240.0)	<.001
PPG area (AU·ms), median (IQR)	15.7 (14.6-16.4)	17.0 (15.9-18.5)	.001

[a]PPG: photoplethysmogram.

[b]AU: arbitrary unit.

Correlations between Hemodynamic Variables and Photoplethysmogram Variables

During the abrupt BP rise because of laryngeal stimulation, changes in the inverse of the PAT showed better correlation with the changes in SBP (r=0.582; 95% CI 0.281-0.779) than with changes in either mean arterial pressure (MAP; r=0.525; 95% CI 0.204-0.745) or diastolic BP (DBP; r=0.442; 95% CI 0.098-0.692). However, these differences in correlation coefficients were not statistically significant when compared with the coefficient between changes in the inverse of PAT and SBP using Fisher Z-transformation (correlation coefficient between changes in the inverse of PAT and MAP, P=.76; between changes in the inverse of PAT and DBP, P=.48).

When we examined the relationship between SBP and PPG-related variables, the changes in the maximum slope of the PPG showed a significant correlation with SBP changes (r=0.49, P=.01). SBP changes also showed a good correlation with changes in the inverse of PAT (r=0.582; P<.001; Figure 3). Other PPG parameters such as amplitude, minimum slope, $width_{50}$, and area did not show a statistically significant correlation with SBP changes (Table 3).

Figure 3. Correlation between the changes in systolic blood pressure and the changes in the inverse of the pulse arrival time (r=0.582, 95% CI 0.281-0.779; *P*<.001).

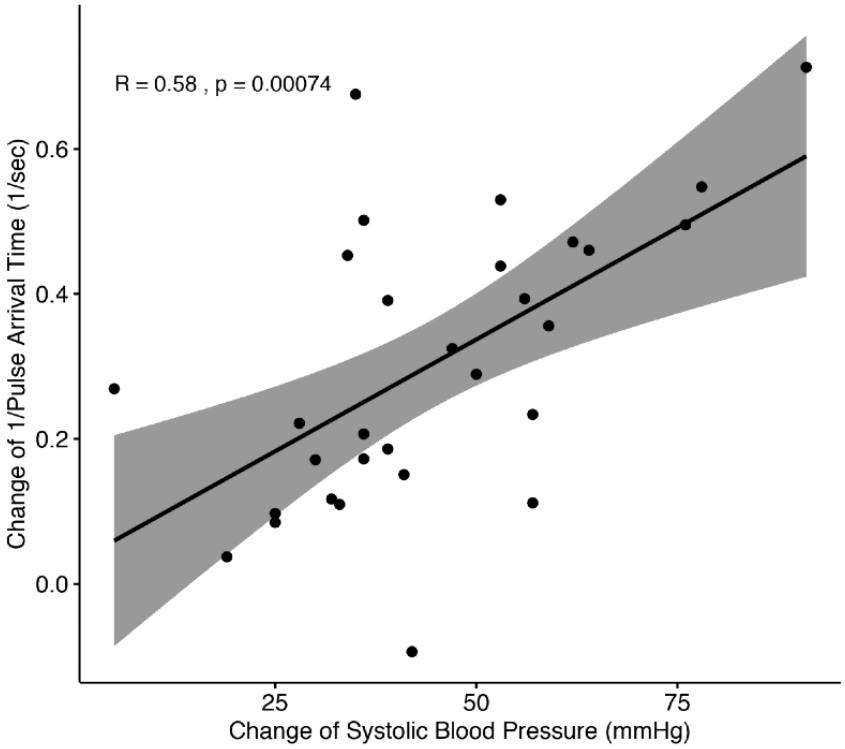


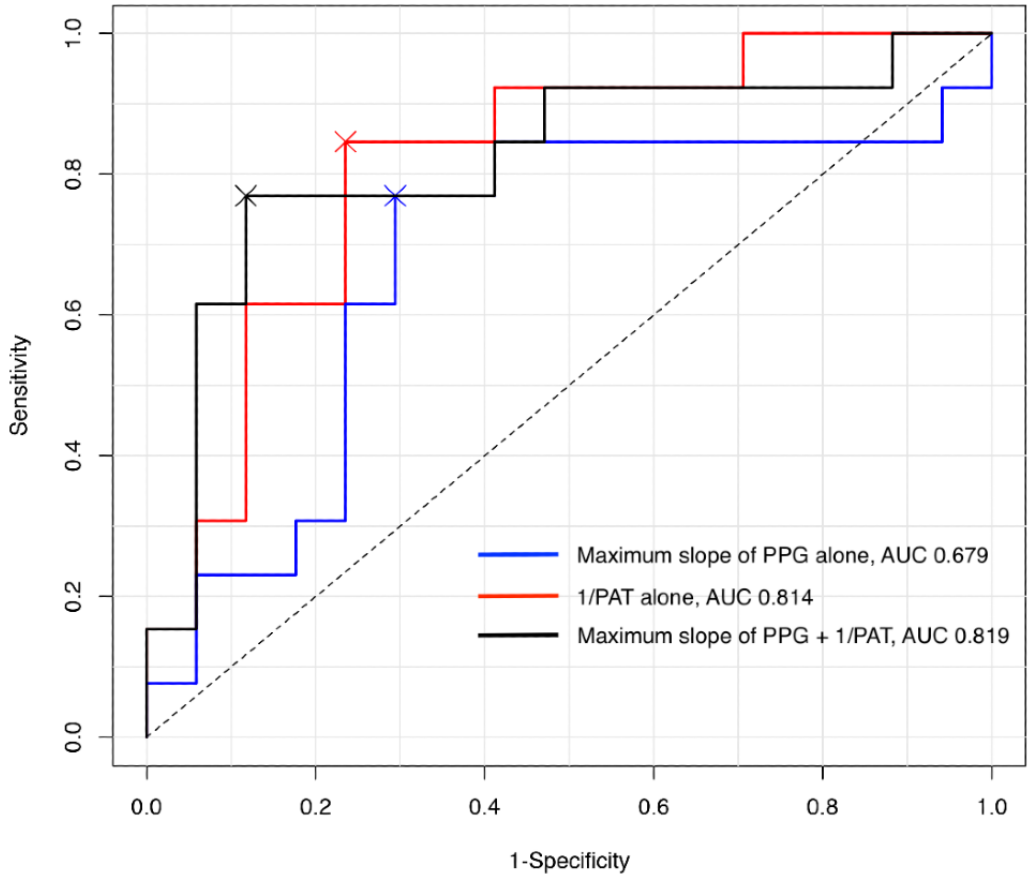
Table 3. Correlation between systolic blood pressure changes and photoplethysmogram-related variables.

Variables	Correlation coefficient, *r*	*P* value
Δ1/pulse arrival time	0.582	<.001
ΔPPG[a] amplitude	0.338	.07
ΔPPG maximum slope	0.468	.009
ΔPPG minimum slope	−0.149	.43
ΔPPG $width_{50}$	0.049	.79
ΔPPG area	0.254	.18

[a]PPG: photoplethysmogram.

PAT and PPG maximal slope showed a significant correlation with the changes in SBP, and in ROC curve analysis, these were evaluated as measurements for detecting greater than 40% change in SBP. In total, 3 models were introduced for the evaluation. The first model detected a 40% change in BP with 1/PAT alone. The second model used the maximum slope of PPG. The last model used both. The model that used changes in the 1/PAT alone to detect 40% or greater changes in the SBP showed an area under the curve (AUC) of 0.814 (95% CI 0.656-0.973 and optimal cutoff value 11.5%), whereas the model that used changes in the maximum slope of PPG showed an AUC of 0.679 (95% CI 0.465-0.892 and optimal cutoff value 11.7%). The combined ROC curve using the changes in both 1/PAT and PPG maximum slope showed an AUC of 0.819 (95% CI 0.650-0.988), which was not significantly different from the model with 1/PAT alone (Delong test, z=−0.114; *P*=.91; Figure 4).

Figure 4. Receiver operating characteristic (ROC) curve analysis of percentage changes in the inverse of the pulse arrival time (PAT) and the maximum slope of photoplethysmogram (PPG) relative to 40% or greater changes in systolic blood pressure. Model using the maximum slope of PPG alone, area under the curve (AUC) 0.679, 95% CI 0.465-0.892, optimal cutoff value 11.7%; model using the inverse of the PAT alone, AUC 0.814, 95% CI 0.656-0.973, optimal cutoff value 11.5%; model using both inverse of the PAT and PPG maximum slope, AUC 0.819, 95% CI 0.650-0.988. Delong test for ROC curves of the model using inverse of PAT alone and the model using both inverse of PAT and PPG maximum slope, z=−0.114; *P*=.91.



Discussion

Principal Findings

We found that beat-to-beat PAT measurements using ECG and PPG are useful noninvasive indicators of BP surge in patients undergoing LMS. Specifically, changes in the inverse of the PAT and systolic NIBP show similar trends, and the inverse of the PAT changes shows a better correlation with the systolic NIBP changes than with those in the MAP or DBP. Moreover, beat-to-beat changes in the PAT reflect the BP surge 1 to 2 min earlier compared with cuff-type NIBP measurements.

Laryngoscopic manipulation in our study is clinically similar to the endotracheal intubation phase in our previous study [11]. PAT did not significantly decrease despite a 24.5% increase in SBP in the previous study. However, this study showed significant changes in SBP and PAT caused by laryngoscopic manipulation. The difference in the intensity of stimulation between LMS and endotracheal intubation might be an explanation for this discrepancy. In addition, the difference in the study population is a possible cause. All subjects of the previous study were hypertensive patients with end-stage renal disease, whereas only 40% of patients in this study had hypertension. It is assumed that the disturbances in autonomic cardiovascular regulation by chronic hypertension can cause these different responses to similar sympathetic stimulation [22].

Although changes in the maximum slope of the PPG showed a correlation with SBP changes, it did not increase the AUC of the ROC curve when it was combined with 1/PAT changes. Rather, the change in the 1/PAT value alone was sufficient to detect a 40% or greater change in the SBP, and the AUC of the ROC curve (AUC 0.814; Figure 4) was comparable with that in our previous study (AUC 0.85) [11]. This suggests that simply combining the features from the PPG signal and PAT does not further increase the diagnostic ability regarding detecting BP changes.

Clinical Applicability of Pulse Arrival Time

LMS is a hemodynamically challenging procedure for anesthesiologists. Laryngoscopic manipulation and suspension of the larynx cause plasma catecholamine levels to rise and induce hemodynamic perturbation, including hypertension, tachycardia, bradycardia, and arrhythmias [7,23-27]. These hemodynamic aberrations can potentially produce poor postoperative outcomes, including prolonged hospital stay and morbidity [28]. As LMS has a short operating time and is usually an outpatient-based procedure, it is not suitable for invasive monitoring such as arterial catheterization unless the patient has known significant preoperative risk factors. However, it is not possible to screen every patient during routine preoperative

evaluation for risks such as undiagnosed intracranial aneurysm. Even in patients without any underlying disease, unexpected changes in BP can be a potential risk for including bleeding, cerebrovascular events, or myocardial ischemia. Hence, simple and noninvasive methods of continuous monitoring are needed for early detection and quick response to unexpected BP change.

During different surgical procedures, various kinds of biosignals can be used to monitor abrupt physiological changes or to predict outcomes. PPG signals, in particular, can be used to reflect various physiological changes such as vascular compliance during hepatic graft reperfusion period following liver transplantation [29] and the success of lumbar sympathetic ganglion block [30]. There have been previous studies on the association between pulse transit time and BP. One study reported that the pulse transit time decreased as the DBP increased because of epinephrine injections in a cohort of 10 dogs [10]. In another study of 10 normal volunteers and 4 hemodialysis patients, PAT was found to correlate with SBP in normal subjects and could detect BP variations in dialysis patients [31]. Ambulatory BP monitoring based on PAT was reported to have significant accuracy in detecting BP above a threshold during nighttime [13]. In a previous evaluation of a monitoring device that estimates BP using a PAT derived by ECG and PPG, the accuracy of the instrument was comparable with that of a monitor based on oscillometry [32]. However, the relationship between BP and PAT is impaired in patients with chronic heart failure, whereas PAT is directly proportional to BP in healthy volunteers [33]. In addition to underlying heart disease, other possible confounding factors, including heart rate, pre-ejection period, and arterial stiffness, should be considered as limitations of BP monitoring using PAT [9,16,34,35]. In an earlier study of patients undergoing maxillofacial surgery, BP estimation using PAT was enhanced when two confounding factors, heart rate and arterial stiffness, were included in the multiple regression analysis model [9]. In this study, one subject showed a PAT increase from 443.4 to 462.6 ms despite a BP surge (systolic NIBP 118 to 160 mmHg; Figure 2). This was thought to be because of a decreased heart rate in this individual (102-75 bpm), which increased with an elevated BP in all other cases. This finding suggests that the refinement of the PAT-based BP-monitoring approach is necessary to reflect confounding factors.

There were some notable findings from this analysis. First, we found that BP changes could be detected 1 to 2 min earlier with PAT than with NIBP, suggesting that PAT monitoring will facilitate earlier responses to BP changes than NIBP monitoring alone. Second, changes in PPG maximum slope also correlated with SBP changes. It is noteworthy that a feature of the PPG waveform alone correlates with BP, regardless of other signals, such as ECG. On the basis of the results of this study, we developed a built-in device, which is able to monitor BP surge using ECG and PPG signal in a continuous and noninvasive manner (Figure 5). Now we are preparing for prospective research to validate the accuracy and to evaluate the usefulness of this device in real clinical practice. We also expect to apply more advanced mathematical techniques to build up a plausible physiologic model of BP estimation. In a recent study that estimated BP from features of the PPG waveform using an artificial neural network, all estimations showed a good correlation with the reference values when the algorithm was validated [36]. Additional research on features of the PPG waveform itself in the estimation of BP is also likely to be of value.

Figure 5. Development of a built-in device to monitor blood pressure surge using electrocardiogram (ECG) and photoplethysmographic (PPG) signal. Tablet personal computer as display and other biosignal sensor system were combined together and enclosed by housing. Blood pressure estimation algorithm was applied to the device, and the software was based on LabVIEW programming. Note that real-time–plotted red dots over ECG and PPG wave on display monitor indicate peak and first derivatives of those signals.

Limitations

Our study has several limitations of note. First, we used PAT measured by ECG R-wave and PPG signal to estimate actual pulse transit time, which should be measured by the time delay between arterial waveforms from two different places. Although the measurement of PAT is noninvasive and is technically easy to perform at low cost, it has been suggested that the PAT may not be a proper surrogate for the correct pulse transit time [37,38]. To calculate the exact pulse transit time, pre-ejection

period should be subtracted from the PAT. However, the pre-ejection period was not measured in this study. Future studies are needed to measure the pre-ejection time, using a method such as the cardiac bioimpedance technique [35], and to investigate its hemodynamic relationships. Second, despite the importance of the heart rate as a confounding factor in the relationship between PAT and BP [9,15,34], adjustment for the heart rate was not performed in this study. The development of techniques to make adjustments for the effect of heart rate on PAT is likely to be meaningful. Third, we used an automated oscillometric NIBP as a routine BP monitor in our patients undergoing LMS. The oscillometric NIBP may not be reliable for measuring actual BP beyond a specific range [39]. However, with advances in technology and manufacturers' efforts to meet medical standards, the accuracy of oscillometric NIBP instruments has been validated [40] and is considered to be reliable in determining BP measurements within the scope of our study. Fourth, instead of performing a calibration of PAT for NIBP, we only assessed the performance of PAT in monitoring relative changes in BP during the intubation period. PAT requires calibration to provide absolute B*P* values because both measurements have different scales and wide ranges of interpatient variability when they are converted to each other. Finally, as we conducted a retrospective analysis, selection bias or unknown confounders cannot be excluded. Well-designed prospective analyses and further device evaluation are warranted in the future.

Conclusions

There is a clear correlation between BP and PAT, which demonstrates the potential of PAT as a useful and noninvasive means of continuous BP monitoring in situations where invasive monitoring is not appropriate. Further studies to adjust for confounding factors are needed to refine the BP-monitoring approach using PAT.

Acknowledgments

The authors acknowledge Drs Hyung-Chul Lee and Chul-Woo Jung (Seoul National University) for their efforts in developing Vital Recorder, which was used to collect and process the biosignal data in this study. This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute, funded by the Ministry of Health and Welfare, Republic of Korea (grant number: HI17C2410), and was also supported by the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Science, ICT and Future Planning (grant Number 2016M3A9E8941259, 2019R1A2C4069504).

Authors' Contributions

YP was responsible for data acquisition, analysis, and drafting the article. SK designed the research plan, revised the manuscript, and organized the study resources. YL, SC, and SK contributed to the critical revision of the article. GH gave important suggestions during data interpretation and manuscript preparation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
An add-on Python code for calculation of pulse arrival time which is included as a filter function in Vital Recorder software.
[TXT File , 3 KB - jmir_v22i5e13156_app1.txt]

Multimedia Appendix 2
Raw data that were analyzed in this study.
[XLSX File (Microsoft Excel File), 14 KB - jmir_v22i5e13156_app2.xlsx]

References

1. Kovac AL. Controlling the hemodynamic response to laryngoscopy and endotracheal intubation. J Clin Anesth 1996 Feb;8(1):63-79. [doi: 10.1016/0952-8180(95)00147-6] [Medline: 8695083]
2. Mustola ST, Baer GA, Metsä-Ketelä T, Laippala P. Haemodynamic and plasma catecholamine responses during total intravenous anaesthesia for laryngomicroscopy. Thiopentone compared with propofol. Anaesthesia 1995 Feb;50(2):108-113 [FREE Full text] [doi: 10.1111/j.1365-2044.1995.tb15090.x] [Medline: 7710018]
3. Shribman AJ, Smith G, Achola KJ. Cardiovascular and catecholamine responses to laryngoscopy with and without tracheal intubation. Br J Anaesth 1987 Mar;59(3):295-299 [FREE Full text] [doi: 10.1093/bja/59.3.295] [Medline: 3828177]
4. Cheung AT. Exploring an optimum intra/postoperative management strategy for acute hypertension in the cardiac surgery patient. J Card Surg 2006;21(Suppl 1):S8-14. [doi: 10.1111/j.1540-8191.2006.00214.x] [Medline: 16492294]
5. Howell SJ, Sear JW, Foëx P. Hypertension, hypertensive heart disease and perioperative cardiac risk. Br J Anaesth 2004 Apr;92(4):570-583 [FREE Full text] [doi: 10.1093/bja/aeh091] [Medline: 15013960]

6. Strong MS, Vaughan CW, Mahler DL, Jaffe DR, Sullivan RC. Cardiac complications of microsurgery of the larynx: etiology, incidence and prevention. Laryngoscope 1974 Jun;84(6):908-920. [doi: 10.1288/00005537-197406000-00003] [Medline: 4832373]
7. Pandazi AK, Louizos AA, Davilis DJ, Stivaktakis JM, Georgiou LG. Inhalational anesthetic technique in microlaryngeal surgery: a comparison between sevoflurane-remifentanil and sevoflurane-alfentanil anesthesia. Ann Otol Rhinol Laryngol 2003 Apr;112(4):373-378. [doi: 10.1177/000348940311200414] [Medline: 12731635]
8. Steinfath M, Scholz J, Tonner PH, Kycia B, Konietzko T, Bause H, et al. Is sufentanil a useful opioid for laryngomicroscopy? Acta Anaesthesiol Scand 1996 Sep;40(8 Pt 1):883-888. [doi: 10.1111/j.1399-6576.1996.tb04554.x] [Medline: 8908222]
9. Baek HJ, Kim KK, Kim JS, Lee B, Park KS. Enhancing the estimation of blood pressure using pulse arrival time and two confounding factors. Physiol Meas 2010 Feb;31(2):145-157. [doi: 10.1088/0967-3334/31/2/002] [Medline: 20009186]
10. Geddes LA, Voelz MH, Babbs CF, Bourland JD, Tacker WA. Pulse transit time as an indicator of arterial blood pressure. Psychophysiology 1981 Jan;18(1):71-74. [doi: 10.1111/j.1469-8986.1981.tb01545.x] [Medline: 7465731]
11. Kim S, Song J, Park J, Kim J, Park Y, Hwang G. Beat-to-beat tracking of systolic blood pressure using noninvasive pulse transit time during anesthesia induction in hypertensive patients. Anesth Analg 2013 Jan;116(1):94-100. [doi: 10.1213/ANE.0b013e318270a6d9] [Medline: 23223109]
12. Sharwood-Smith G, Bruce J, Drummond G. Assessment of pulse transit time to indicate cardiovascular changes during obstetric spinal anaesthesia. Br J Anaesth 2006 Jan;96(1):100-105 [FREE Full text] [doi: 10.1093/bja/aei266] [Medline: 16257996]
13. Zheng Y, Poon CC, Yan BP, Lau JY. Pulse arrival time based cuff-less and 24-H wearable blood pressure monitoring and its diagnostic value in hypertension. J Med Syst 2016 Sep;40(9):195. [doi: 10.1007/s10916-016-0558-6] [Medline: 27447469]
14. Gesche H, Grosskurth D, Küchler G, Patzak A. Continuous blood pressure measurement by using the pulse transit time: comparison to a cuff-based method. Eur J Appl Physiol 2012 Jan;112(1):309-315. [doi: 10.1007/s00421-011-1983-3] [Medline: 21556814]
15. Wang R, Jia W, Mao Z, Sclabassi RJ, Sun M. Cuff-free blood pressure estimation using pulse transit time and heart rate. Int Conf Signal Process Proc 2014 Oct;2014:115-118 [FREE Full text] [doi: 10.1109/ICOSP.2014.7014980] [Medline: 26213717]
16. Wong MY, Pickwell-MacPherson E, Zhang YT, Cheng JC. The effects of pre-ejection period on post-exercise systolic blood pressure estimation using the pulse arrival time technique. Eur J Appl Physiol 2011 Jan;111(1):135-144. [doi: 10.1007/s00421-010-1626-0] [Medline: 20824282]
17. Carek AM, Conant J, Joshi A, Kang H, Inan OT. SeismoWatch: wearable cuffless blood pressure monitoring using pulse transit time. Proc ACM Interact Mob Wearable Ubiquitous Technol 2017 Sep;1(3):pii: 40 [FREE Full text] [doi: 10.1145/3130905] [Medline: 30556049]
18. Liang Y, Abbott D, Howard N, Lim K, Ward R, Elgendi M. How effective is pulse arrival time for evaluating blood pressure? Challenges and recommendations from a study using the MIMIC database. J Clin Med 2019 Mar 11;8(3):pii: E337 [FREE Full text] [doi: 10.3390/jcm8030337] [Medline: 30862031]
19. Lee H, Jung C. Vital Recorder-a free research tool for automatic recording of high-resolution time-synchronised physiological data from multiple anaesthesia devices. Sci Rep 2018 Jan 24;8(1):1527 [FREE Full text] [doi: 10.1038/s41598-018-20062-4] [Medline: 29367620]
20. Kemp B, Värri A, Rosa AC, Nielsen KD, Gade J. A simple format for exchange of digitized polygraphic recordings. Electroencephalogr Clin Neurophysiol 1992 May;82(5):391-393. [doi: 10.1016/0013-4694(92)90009-7] [Medline: 1374708]
21. Alian AA, Galante NJ, Stachenfeld NS, Silverman DG, Shelley KH. Impact of central hypovolemia on photoplethysmographic waveform parameters in healthy volunteers. Part 1: time domain analysis. J Clin Monit Comput 2011 Dec;25(6):377-385. [doi: 10.1007/s10877-011-9316-y] [Medline: 22051898]
22. Mancia G, Grassi G. The autonomic nervous system and hypertension. Circ Res 2014 May 23;114(11):1804-1814. [doi: 10.1161/CIRCRESAHA.114.302524] [Medline: 24855203]
23. Ayuso A, Luis M, Sala X, Sánchez J, Traserra J. Effects of anesthetic technique on the hemodynamic response to microlaryngeal surgery. Ann Otol Rhinol Laryngol 1997 Oct;106(10 Pt 1):863-868. [doi: 10.1177/000348949710601010] [Medline: 9342984]
24. Durrani M, Barwise JA, Johnson RF, Kambam JR, Janicki PK. Intravenous chloroprocaine attenuates hemodynamic changes associated with direct laryngoscopy and tracheal intubation. Anesth Analg 2000 May;90(5):1208-1212. [doi: 10.1097/00000539-200005000-00039] [Medline: 10781481]
25. Louizos AA, Hadzilia SJ, Davilis DI, Samanta EG, Georgiou LG. Administration of esmolol in microlaryngeal surgery for blunting the hemodynamic response during laryngoscopy and tracheal intubation in cigarette smokers. Ann Otol Rhinol Laryngol 2007 Feb;116(2):107-111. [doi: 10.1177/000348940711600205] [Medline: 17388233]
26. Lee CW, Kim M. Effects of preanesthetic dexmedetomidine on hemodynamic responses to endotracheal intubation in elderly patients undergoing treatment for hypertension: a randomized, double-blinded trial. Korean J Anesthesiol 2017 Feb;70(1):39-45 [FREE Full text] [doi: 10.4097/kjae.2017.70.1.39] [Medline: 28184265]

27. Ko Y, Kim Y, Park S, Chung WS, Noh C, Lee J. Comparison of etomidate and propofol on intubating conditions and the onset time associated with cisatracurium administration. Korean J Anesthesiol 2015 Apr;68(2):136-140 [FREE Full text] [doi: 10.4097/kjae.2015.68.2.136] [Medline: 25844131]
28. Reich DL, Bennett-Guerrero E, Bodian CA, Hossain S, Winfree W, Krol M. Intraoperative tachycardia and hypertension are independently associated with adverse outcome in noncardiac surgery of long duration. Anesth Analg 2002 Aug;95(2):273-7, table of contents. [doi: 10.1097/00000539-200208000-00003] [Medline: 12145033]
29. Kim W, Kim J, Moon Y, Kim S, Hwang G, Shin W. The photoplethysmographic amplitude to pulse pressure ratio can track sudden changes in vascular compliance and resistance during liver graft reperfusion: A beat-to-beat analysis. Medicine (Baltimore) 2017 Jun;96(22):e7045 [FREE Full text] [doi: 10.1097/MD.0000000000007045] [Medline: 28562562]
30. Joo E, Kong YG, Lee J, Cho H, Kim S, Suh JH. Change in pulse transit time in the lower extremity after lumbar sympathetic ganglion block: an early indicator of successful block. J Int Med Res 2017 Feb;45(1):203-210 [FREE Full text] [doi: 10.1177/0300060516681398] [Medline: 28222636]
31. Ahlstrom C, Johansson A, Uhlin F, Länne T, Ask P. Noninvasive investigation of blood pressure changes using the pulse wave transit time: a novel approach in the monitoring of hemodialysis patients. J Artif Organs 2005;8(3):192-197. [doi: 10.1007/s10047-005-0301-4] [Medline: 16235036]
32. Heard SO, Lisbon A, Toth I, Ramasubramanian R. An evaluation of a new continuous blood pressure monitoring system in critically ill patients. J Clin Anesth 2000 Nov;12(7):509-518. [doi: 10.1016/s0952-8180(00)00201-4] [Medline: 11137411]
33. Wagner DR, Roesch N, Harpes P, Körtke H, Plumer P, Saberin A, et al. Relationship between pulse transit time and blood pressure is impaired in patients with chronic heart failure. Clin Res Cardiol 2010 Oct;99(10):657-664. [doi: 10.1007/s00392-010-0168-0] [Medline: 20473677]
34. Lantelme P, Mestre C, Lievre M, Gressard A, Milon H. Heart rate: an important confounder of pulse wave velocity assessment. Hypertension 2002 Jun;39(6):1083-1087. [doi: 10.1161/01.hyp.0000019132.41066.95] [Medline: 12052846]
35. Payne RA, Symeonides CN, Webb DJ, Maxwell SR. Pulse transit time measured from the ECG: an unreliable marker of beat-to-beat blood pressure. J Appl Physiol (1985) 2006 Jan;100(1):136-141 [FREE Full text] [doi: 10.1152/japplphysiol.00657.2005] [Medline: 16141378]
36. Xing X, Sun M. Optical blood pressure estimation with photoplethysmography and FFT-based neural networks. Biomed Opt Express 2016 Aug 1;7(8):3007-3020 [FREE Full text] [doi: 10.1364/BOE.7.003007] [Medline: 27570693]
37. Naschitz JE, Bezobchuk S, Mussafia-Priselac R, Sundick S, Dreyfuss D, Khorshidi I, et al. Pulse transit time by R-wave-gated infrared photoplethysmography: review of the literature and personal experience. J Clin Monit Comput 2004 Dec;18(5-6):333-342. [doi: 10.1007/s10877-005-4300-z] [Medline: 15957624]
38. Zhang G, Gao M, Xu D, Olivier NB, Mukkamala R. Pulse arrival time is not an adequate surrogate for pulse transit time as a marker of blood pressure. J Appl Physiol (1985) 2011 Dec;111(6):1681-1686 [FREE Full text] [doi: 10.1152/japplphysiol.00980.2011] [Medline: 21960657]
39. Ribezzo S, Spina E, Di Bartolomeo S, Sanson G. Noninvasive techniques for blood pressure measurement are not a reliable alternative to direct measurement: a randomized crossover trial in ICU. ScientificWorldJournal 2014;2014:353628 [FREE Full text] [doi: 10.1155/2014/353628] [Medline: 24616624]
40. O'Brien E, Stergiou GS. The pursuit of accurate blood pressure measurement: A 35-year travail. J Clin Hypertens (Greenwich) 2017 Aug;19(8):746-752 [FREE Full text] [doi: 10.1111/jch.13005] [Medline: 28429879]

Abbreviations

AUC: area under the curve
BP: blood pressure
DBP: diastolic blood pressure
ECG: electrocardiogram
LMS: laryngeal microsurgery
MAP: mean arterial pressure
NIBP: noninvasive intermittent blood pressure
PAT: pulse arrival time
PPG: photoplethysmogram
ROC: receiver operating characteristic
SBP: systolic blood pressure

Edited by G Eysenbach; submitted 15.12.18; peer-reviewed by M Lang, S Mousavi; comments to author 27.04.19; revised version received 22.06.19; accepted 09.02.20; published 15.05.20.

Please cite as:
Park YS, Kim SH, Lee YS, Choi SH, Ku SW, Hwang GS
Real-Time Monitoring of Blood Pressure Using Digitalized Pulse Arrival Time Calculation Technology for Prompt Detection of Sudden Hypertensive Episodes During Laryngeal Microsurgery: Retrospective Observational Study
J Med Internet Res 2020;22(5):e13156
URL: https://www.jmir.org/2020/5/e13156
doi:10.2196/13156
PMID:32412413

Original Paper

Using Biosensors and Digital Biomarkers to Assess Response to Cardiac Rehabilitation: Observational Study

Hélène De Cannière[1,2], MSc; Christophe J P Smeets[1,2,3], PhD; Melanie Schoutteten[1,2], MD; Carolina Varon[4,5], PhD; Chris Van Hoof[4,6], PhD; Sabine Van Huffel[4], PhD; Willemijn Groenendaal[3], PhD; Pieter Vandervoort[1,2,7], MD

[1]Mobile Health Unit, Limburg Clinical Research Center (LCRC), Faculty of Medicine and Life Sciences, Hasselt University (UHasselt), Diepenbeek, Belgium

[2]Department of Future Health, Ziekenhuis Oost-Limburg, Genk, Belgium

[3]Holst Centre, imec the Netherlands, Eindhoven, Netherlands

[4]Center for Dynamical Systems, Signal Processing and Data Analytics (STADIUS), Department of Electrical Engineering (ESAT), KU (Katholieke Universiteit) Leuven, Leuven, Belgium

[5]Circuits and Systems (CAS), Department of Microelectronics, Delft University of Technology (TU Delft), Delft, Netherlands

[6]imec vzw Belgium, Leuven, Belgium

[7]Department of Cardiology, Ziekenhuis Oost-Limburg, Genk, Belgium

Corresponding Author:
Hélène De Cannière, MSc
Mobile Health Unit, Limburg Clinical Research Center (LCRC)
Faculty of Medicine and Life Sciences
Hasselt University (UHasselt)
Agoralaan
Diepenbeek, 3590
Belgium
Phone: 32 474811834
Email: helene.decanniere@uhasselt.be

Abstract

Background: Cardiac rehabilitation (CR) is known for its beneficial effects on functional capacity and is a key component within current cardiovascular disease management strategies. In addition, a larger increase in functional capacity is accompanied by better clinical outcomes. However, not all patients respond in a similar way to CR. Therefore, a patient-tailored approach to CR could open up the possibility to achieve an optimal increase in functional capacity in every patient. Before treatment can be optimized, the differences in response of patients in terms of cardiac adaptation to exercise should first be understood. In addition, digital biomarkers to steer CR need to be identified.

Objective: The aim of the study was to investigate the difference in cardiac response between patients characterized by a clear improvement in functional capacity and patients showing only a minor improvement following CR therapy.

Methods: A total of 129 patients in CR performed a 6-minute walking test (6MWT) at baseline and during four consecutive short-term follow-up tests while being equipped with a wearable electrocardiogram (ECG) device. The 6MWTs were used to evaluate functional capacity. Patients were divided into high- and low-response groups, based on the improvement in functional capacity during the CR program. Commonly used heart rate parameters and cardiac digital biomarkers representative of the heart rate behavior during the 6MWT and their evolution over time were investigated.

Results: All participating patients improved in functional capacity throughout the CR program (P<.001). The heart rate parameters, which are commonly used in practice, evolved differently for both groups throughout CR. The peak heart rate (HR_{peak}) from patients in the high-response group increased significantly throughout CR, while no change was observed in the low-response group ($F_{4,92}$=8.321, P<.001). Similar results were obtained for the recovery heart rate (HR_{rec}) values, which increased significantly over time during every minute of recuperation, for the high-response group (HR_{rec1}: P<.001, HR_{rec2}: P<.001, HR_{rec3}: P<.001, HR_{rec4}: P<.001, and HR_{rec5}: P=.02). The other digital biomarkers showed that the evolution of heart rate behavior during a standardized activity test differed throughout CR between both groups. These digital biomarkers, derived from the continuous measurements, contribute to more in-depth insight into the progression of patients' cardiac responses.

Conclusions: This study showed that when using wearable sensor technology, the differences in response of patients to CR can be characterized by means of commonly used heart rate parameters and digital biomarkers that are representative of cardiac response to exercise. These digital biomarkers, derived by innovative analysis techniques, allow for more in-depth insights into the cardiac response of cardiac patients during standardized activity. These results open up the possibility to optimized and more patient-tailored treatment strategies and to potentially improve CR outcome.

(J Med Internet Res 2020;22(5):e17326) doi:10.2196/17326

KEYWORDS

wearables; sensor; 6MWT; rehabilitation; cardiovascular

Introduction

Cardiovascular diseases are the most prevalent noncommunicable diseases worldwide. The American College of Cardiology Foundation, the American Heart Association, and the European Society of Cardiology consider cardiac rehabilitation (CR) to be a key component within current disease management strategies, making millions of cardiac patients eligible for rehabilitation [1-3]. Unfortunately, CR programs worldwide are characterized by low implementation rates, as 33%-71% of eligible patients are not referred [4]. Notwithstanding, participation in a CR program has shown to increase cardiorespiratory fitness, thereby improving physiological responses to physical effort [5]. The improvement in functional capacity is clinically relevant as it not only improves quality of life in patients, but also serves as a powerful predictor for mortality [6]. These benefits of CR on mortality, morbidity, and quality of life have been studied comprehensively in several meta-analyses [7-10]. Moreover, the beneficial effects of CR seem to be even more pronounced when functional capacity is increased to a larger extent [11-14]. Although previous research showed that exercise-based CR has beneficial and clinically relevant effects on functional capacity, a large variability on the response to training is seen among CR patients. Recent studies have shown that *responders* are characterized by lower baseline peak VO_2 (peak oxygen uptake) values and a reduced baseline ejection fraction, while *nonresponders* have an impaired chronotropic competence, which predicts poor training response [15-17]. Other potential contributing factors to poor training response include adherence rates, exercise dose, functioning of the autonomic system, or comorbidities [18]. Nevertheless, little is known about the mechanisms causing the large variability. Therefore, Gevaert et al stated that future research needs to focus on studying these contributing factors in order to generate the best response to CR [18]. Moreover, research focusing on changes occurring during CR, and not only before or after completing the program, can contribute in the development of a more patient-tailored CR program.

The aim of this study was to investigate the difference in cardiac response, a measure of chronotropic response, between patients that showed a clear improvement in functional capacity and patients that only showed a minor improvement following CR therapy. This was done by using data captured with a wearable electrocardiogram (ECG) device during a standardized activity. Moreover, innovative analysis techniques were used to derive digital cardiac biomarkers allowing an in-depth analysis of heart rate behavior during a standardized activity test.

Methods

Study Design

A total of 129 cardiovascular patients, who were enrolled in a multidisciplinary CR program in a single tertiary-care center (Ziekenhuis Oost-Limburg [ZOL], Genk, Belgium) and representative of the typical CR population, were included. Patients over the age of 18 years with heart failure and reduced ejection fraction, with heart failure and preserved ejection fraction, and with a left ventricular ejection fraction less than or equal to 55% were eligible for the study. Patients with an inability to exercise due to orthopedic or neurological limitations were excluded from the study. The goal was to investigate the different levels of response to exercise intervention during a standardized CR program. The 6-minute walking test (6MWT) was used to follow up on the improvement in functional capacity in the course of the CR program. A wearable device was used to collect ECG data during the 6MWT. A descriptive analysis of the longitudinally collected wearable data was performed to identify patterns or trends in the dataset. To distinguish the response to rehabilitation, patients were assessed as being within a low-response and a high-response group based on a median split for the increase in distance walked throughout the CR program. Therefore, two groups with an equal number of patients were created based on the level of improvement in functional capacity measured by the 6-minute walking distance. Patients who increased more than 90 meters after completing the CR were referred to as the high-response group, while the low-response group consisted of patients who increased less than 90 meters. The study complied with the Declaration of Helsinki, and the local ethical committee approved the study protocol. All subjects gave written informed consent prior to study participation.

Multidisciplinary Cardiac Rehabilitation Program

Patients were referred to the multidisciplinary CR program following a cardiovascular-related hospital admission. The 15-week program consisted of 45 ambulatory rehabilitation sessions at a frequency of three 1-hour sessions per week. Both resistive and aerobic exercises were included in the program. Additionally, dietary sessions, psychological support, and social consultations were included in the multidisciplinary program. By standard, a cardiopulmonary exercise test (CPET) was performed at baseline and at end-of-study to assess functional capacity. A total of 14 low-response group patients out of 45 (31%) and 21 high-response group patients out of 45 (47%) had a CPET at both baseline and end-of-study. The heart rate achieved at 90% of ventilator threshold during the CPET was

chosen as the target heart rate during aerobic training. If no CPET data were available, target heart rate was set at 50%-80% of the maximal heart rate. Aerobic training consisted of 30-40 minutes, in total, of aerobic exercise on bicycle, hand bike, treadmill, and/or stepper. Resistive training was performed at 50%-80% of one repetition maximum and consisted of three sets of 15 repetitions on both the leg and arm press. Training intensity was increased every 2 weeks based on patient improvement according to the standard clinical practice of CR in our study center.

Experimental Protocol and Sensor Technology

Demographics, clinical data, medical therapy, and echocardiography data were collected from the electronic medical record. A 6MWT was performed at baseline (ie, start of rehabilitation program). Four follow-up 6MWTs were performed every 3 weeks, resulting in five 6MWTs in total. The compliance rate with the rehabilitation program between every 6MWT for both groups was calculated. Patients were expected to follow three rehabilitation sessions per week, which is equal to nine rehabilitation sessions between two consecutive 6MWT measurements. If patients attended nine rehabilitation sessions between consecutive 6MWTs, a compliance rate of 100% was obtained. The 6MWT was performed according to a standardized protocol [19]. The distance walked after 6 minutes was recorded and was used to check functional capacity during CR. During the 6MWT, all enrolled patients were equipped with a wearable device. The wearable device was equipped with the MUlti SEnsor IC (MUSEIC) chip, supporting a wide range of sensor modalities, including ECG (512 Hz sample frequency) and accelerometer data (32 Hz sample frequency) (imec the Netherlands, Eindhoven, the Netherlands) [20]. The electrodes were positioned according to lead II of Einthoven's triangle [21]. Prior to the 6MWT, patients were at rest for 5 minutes to record their resting heart rate. Additionally, a recuperation phase of 5 minutes was included after the 6MWT to record recovery heart rate.

Preprocessing and Calculation of Static and Dynamic Heart Rate Parameters

The signal was divided into three parts: a 5-minute resting phase, a 6-minute walking phase, and a 5-minute recuperation phase. First, the artefacts present in the ECG signals were automatically detected and removed by means of the algorithm proposed by Varon et al [22]. Next, an algorithm performed the initial automatic R-peak detection [23] and incorrect detections were visually corrected. The R-peaks were then used to generate the heart rate by dividing the signal into 2-minute or 16-second windows with a 4-second stride, from which the heart rate parameters were derived. The heart rate parameters derived from the 2-minute windows were used for further analysis. A distinction was made between two types of these digital cardiac biomarkers. The static, commonly used heart rate parameters that comprise the information of specific periods of time during a standardized activity are placed into a single output parameter, whereas the other digital cardiac biomarkers represent the evolution of heart rate throughout the entire time span of a standardized activity test.

The resting heart rate (HR_{rest}) was calculated by taking the mean heart rate during the final 20 seconds of the resting period. The peak heart rate (HR_{peak}) was calculated by taking the mean heart rate obtained during the final 10 seconds of the walking phase. For the recovery heart rate (HR_{rec}), the mean heart rate during every minute of recuperation following the 6MWT (ie, HR_{rec1}, HR_{rec2}, HR_{rec3}, HR_{rec4}, and HR_{rec5}) was calculated. Moreover, to study whether HR_{peak} was influenced by the difference in effort among patients, HR_{peak} was corrected for the distance walked by dividing HR_{peak} by distance ($HR_{peak\text{-}dist}$), as described previously [24]. These heart rate parameters, commonly used in practice, were calculated for each measurement session separately. The accelerometer data has been used to estimate the effort during the walking phase of the 6MWT. Effort has been previously used as a measure of physical activity intensity [25-27]. The effort has been calculated for the full 6-minute-test duration using the following formula:

in which n is the total number of accelerometer sample points considered and X_k is a vector representing the acceleration along the x-axis, while the other axes are represented by Y_k and Z_k vectors, respectively.

Four different models were used to study the heart rate behavior during a standardized activity for both patient subgroups: one-term, two-term, two-term with added coefficient, and quadratic polynomial models. The goodness of fit was determined by calculating the coefficient of determination (R-squared). The model with the best fit was used to study the heart rate behavior. The coefficients of the best fits—the digital cardiac biomarkers—were studied for differences between both groups.

Study Endpoints

This study focused on investigating the difference in cardiac response, reflected by changes in commonly used heart rate parameters and digital cardiac biomarkers, between patients with a clear increased functional capacity and patients with only a minor improvement after completing CR.

Statistics

Continuous variables are expressed as mean (SD), if normally distributed, or as median (IQR), if nonnormally distributed, and dichotomous data are expressed as n (%). Normality was checked by the Shapiro-Wilk statistic. Categorical data were expressed as numbers and percentages and compared with the Fisher exact test. Continuous variables were compared between groups with the Student *t* test or the Mann-Whitney U test as appropriate. A two-way mixed analysis of variance (ANOVA) investigated the effect of magnitude of improvement on the progression of heart rate measures throughout a CR program. Results are expressed as df_{main}, df_{error}, F, and partial η^2. df_{main} indicates degrees of freedom for the simple main effect and df_{error} indicates degrees of freedom for the error term. F indicates that we are comparing results to an F distribution and partial η^2 is a measure of effect size. If significant, a repeated-measures

ANOVA was performed to analyze the simple main effect over time and univariate analysis was performed to analyze the main effect of response groups. An independent *t* test was performed to compare cardiac biomarkers between groups at specific moments in time. Outliers were assessed by inspection of a boxplot for values greater than 3 box-lengths from the edge of the box and were removed from analysis. The statistical significance was always set at a two-tailed probability level of <.05. Statistics were performed using SPSS version 24.0 (IBM Corp).

Results

Demographics and Baseline Population

Of the 129 patients that consented to participate, 89 (69.0%) completed the total study protocol. Out of 129 patients, 40 (31.0%) were excluded from analysis upon failure to complete the CR program due to health-related problems, lack of motivation, and work or family commitment (see Multimedia Appendix 1). The 89 patients who completed the study protocol were subdivided into two groups based on their improvement in functional capacity throughout the CR. The high-response group consisted of 45 patients who improved more than 90 meters throughout the CR, while the low-response group consisted of 44 patients who improved less than 90 meters. Baseline characteristics of both groups are provided in Table 1. There was no statistical difference between the groups with respect to the demographics and baseline characteristics, except for diabetes. More often, patients in the low-response group suffered from diabetes compared to the high-response group (24.4% vs 2.3%, *P*<.01). The two-way mixed ANOVA showed a similar compliance rate between both subgroups throughout the rehabilitation program ($F_{3,255}$=1.03, *P*=.99). The low- and high-response groups showed a similar average compliance rate of 88.0% and 85.8%, respectively, during the initial 3 weeks of the program, which did not significantly change throughout the remaining 12 weeks of the program.

Functional Capacity

Patients showed an increase in functional capacity based on both the results of the CPET measurements and the results of the 6MWTs. Only 14 patients out of 44 in the low-response group (32%) and 21 patients out of 45 in the high-response group (47%) performed both a CPET at baseline and at end-of-study. Although not statistically significant, an increase of 2.23 mL kg^{-1} min^{-1} in peak VO_2 was seen between baseline and end-of-study for the low-response group (14/44, 32%, *P*=.10). The same group showed a mean significant increase of 55 meters in 6MWT distance between baseline and end-of-study (*P*<.001). The high-response group showed a significant increase of 5.10 mL kg^{-1} min^{-1} in peak VO_2 (21/45, 47%, *P*<.001) and 147 meters in 6MWT distance between baseline and end-of-study (*P*<.001).

Table 1. Baseline characteristics.

Variable	Low-response group[a] (n=45)	High-response group[b] (n=44)	*P* value
Demographics			
Gender (male), n (%)	30 (67)	35 (80)	.23
Age (years), mean (SD)	64 (9)	63 (10)	.83
Height (m), mean (SD)	1.72 (0.09)	1.73 (0.09)	.44
Body surface area (m^2), mean (SD)	1.96 (0.19)	1.93 (0.19)	.45
Active smoker, n (%)	12 (27)	6 (14)	.19
Left ventricle ejection fraction (%), mean (SD)	47 (12)	44 (14)	.18
Cardiac resynchronization therapy, n (%)	2 (4)	2 (5)	>.99
Reason for referral, n (%)			
Myocardial infarction	14 (31)	9 (20)	.33
Heart failure	11 (24)	10 (23)	>.99
Coronary artery bypass grafting	6 (13)	8 (18)	.57
Percutaneous coronary intervention	4 (9)	2 (5)	.68
Comorbidities, n (%)			
Atrial fibrillation	10 (22)	12 (27)	.63
Hypertension	16 (36)	22 (50)	.20
Dyslipidemia	20 (44)	19 (43)	>.99
Diabetes	11 (24)	1 (2)	.004
New York Heart Association class, n (%)			
Class I	11 (24)	14 (32)	.62
Class II	24 (53)	20 (45)	
Class III	9 (20)	10 (23)	
Medications, n (%)			
Angiotensin converting enzyme inhibitor	25 (56)	25 (57)	>.99
Beta-blocker	33 (73)	32 (73)	>.99
Diuretics	19 (42)	20 (45)	.83
Baseline CPET[c] peak VO_2[d] (mL/kg·min), mean (SD)	17.2 (5.3)	16.8 (4.9)	.72
Baseline 6MWT[e] distance (m), mean (SD)	496 (95)	473 (97)	.25
Compliance rate (%), mean (SD)			.99
Baseline to first measurement	88.0 (17.0)	85.5 (13.9)	
First to second measurement	87.7 (18.5)	85.7 (14.1)	
Second to third measurement	84.5 (17.9)	81.9 (20.6)	
Third to end-of-study measurement	85.4 (21.8)	83.5 (24.9)	

[a]This group consisted of patients who improved less than 90 meters throughout the cardiac rehabilitation.

[b]This group consisted of patients who improved more than 90 meters throughout the cardiac rehabilitation.

[c]CPET: cardiopulmonary exercise test.

[d]VO_2: peak oxygen uptake.

[e]6MWT: 6-minute walking test.

Commonly Used Static Heart Rate Parameters

To study the difference in cardiac response between the two subgroups, commonly used heart rate parameters were derived from the ECG data. Hereto, HR_{rest}, HR_{peak}, and HR_{rec} were analyzed. For these parameters, the differences between the two subgroups for every session and the differences in progression throughout CR were studied. Figure 1 shows the evolution of HR_{rest} throughout CR for both groups. A decreasing trend of

HR_{rest} in the high-response group is observed. In addition, ANOVA did not show a difference in the evolution of HR_{rest} throughout the CR program. Nor was a difference seen in HR_{rest} between the high- and low-response groups at any point in time.

Secondly, the evolution in HR_{peak} across five sessions and, more specifically, the difference between both subgroups was investigated. Figure 2 shows the change in HR_{peak} for both the high- and low-response groups throughout CR.

Figure 1. Resting heart rate (HR_{rest}) for each group throughout cardiac rehabilitation.

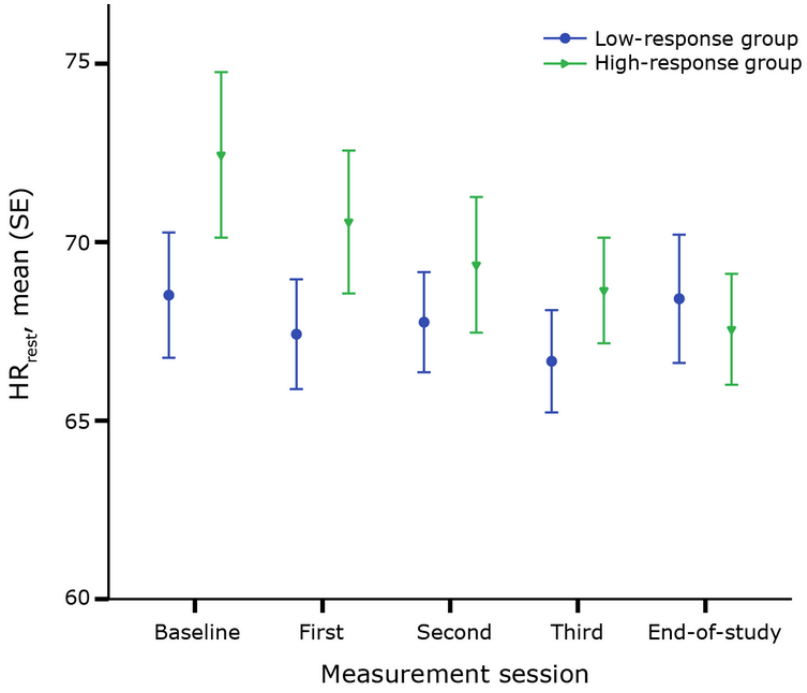
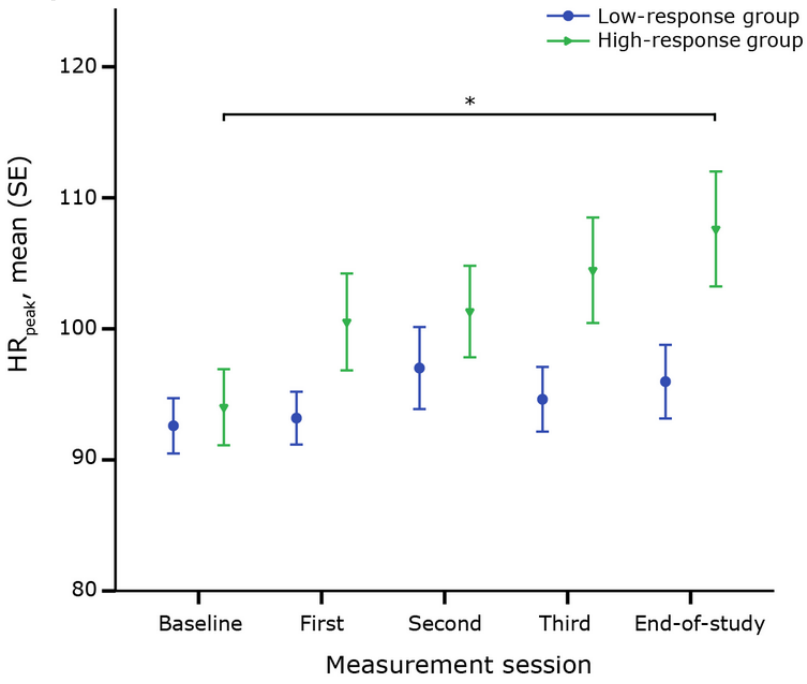


Figure 2. Maximum heart rate (HR_{peak}) for each group throughout cardiac rehabilitation. *denotes a significant change over time.



The two-way mixed ANOVA showed that the evolution of HR_{peak} throughout CR differed between the high- and low-response groups ($F_{4,216}$=3.3, P=.01, partial η^2=.058). HR_{peak} of the high-response group increased significantly throughout rehabilitation, while HR_{peak} of the low-response group remained approximately the same ($F_{4,92}$=8.321, P<.001). Although the evolution differed between both groups, no difference in HR_{peak} was seen at any point in time. HR_{peak} was corrected for distance to study whether the effort shown by patients during the 6MWTs had an influence on HR_{peak}. In Figure 3, both subgroups show a significant decrease in $HR_{peak\text{-}dist}$. However, the decrease in $HR_{peak\text{-}dist}$ was larger for the high-response group compared to

the low-response group. Additionally, a summary measure of the accelerometer data was used to compare the effort between the low- and high-response groups. Both subgroups showed a similar increase in effort throughout rehabilitation and no difference in effort between the groups was seen at any point in time ($F_{4,344}$=.668, *P*=.61).

The change in HR_{rec1} after the 6MWT throughout CR is shown in Figure 4. The evolution of HR_{rec} throughout CR differed among the response groups according to ANOVA for every minute of recuperation (see Table 2).

Figure 3. Peak heart rate corrected for distance ($HR_{peak-dist}$) for each group throughout cardiac rehabilitation. *denotes a significant change over time.

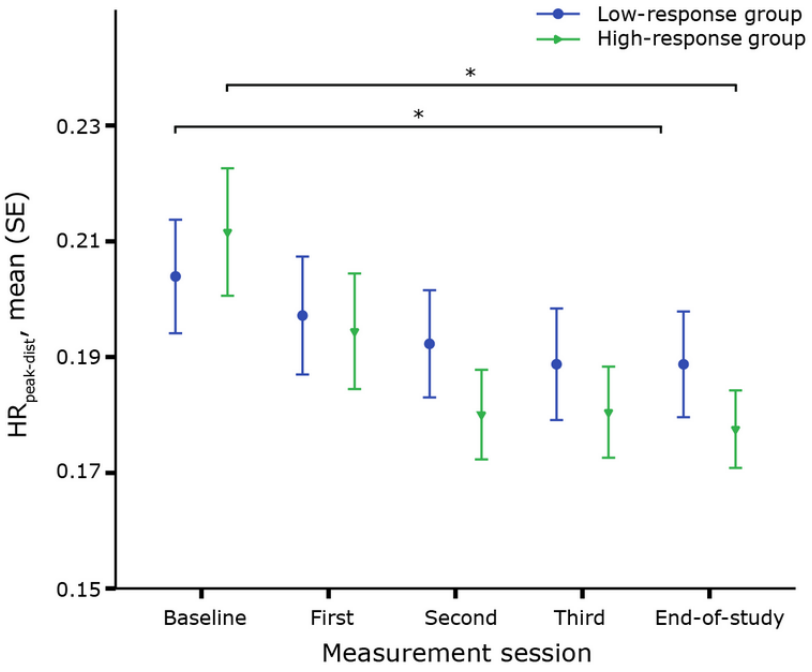


Figure 4. Heart rate recovery during the first minute (HR_{rec1}) after the 6-minute walking test for each group throughout cardiac rehabilitation. *denotes a significant change over time; **denotes a significant difference in HR_{rec1} during a measurement session between both groups.

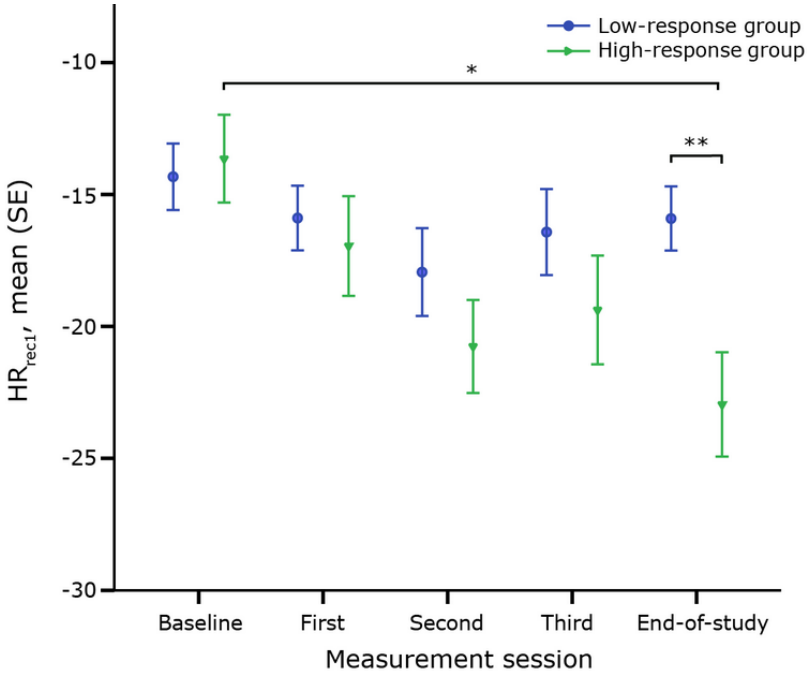


Table 2. Results from the two-way, mixed-model, analysis of variance (ANOVA) for the heart rate recovery (HRrec) during the first 5 minutes after the 6-minute walking test (6MWT).

HR_{rec}[a]	df_{main}[b]	df_{error}[c]	F[d]	Partial η^2[e]	*P* value[f]	*P* value[g]
HR_{rec1}	4	212	5.172	.089	.001	<.001
HR_{rec2}	4	212	7.288	.121	<.001	<.001
HR_{rec3}	4	212	6.634	.111	<.001	<.001
HR_{rec4}	4	212	6.092	.103	<.001	<.001
HR_{rec5}	4	208	3.967	.071	.03	.002

[a]HR_{rec}: heart rate recovery; each number in this column represents every minute of recuperation following the 6MWT.

[b]df_{main}: degrees of freedom for the simple main effect.

[c]df_{error}: degrees of freedom for the error term.

[d]Indicates that we are comparing to an F distribution.

[e]Partial η^2: a measure of effect size.

[f]Significance level for the hypothesis of no time effect × group effect.

[g]Significance level for the hypothesis of no time effect.

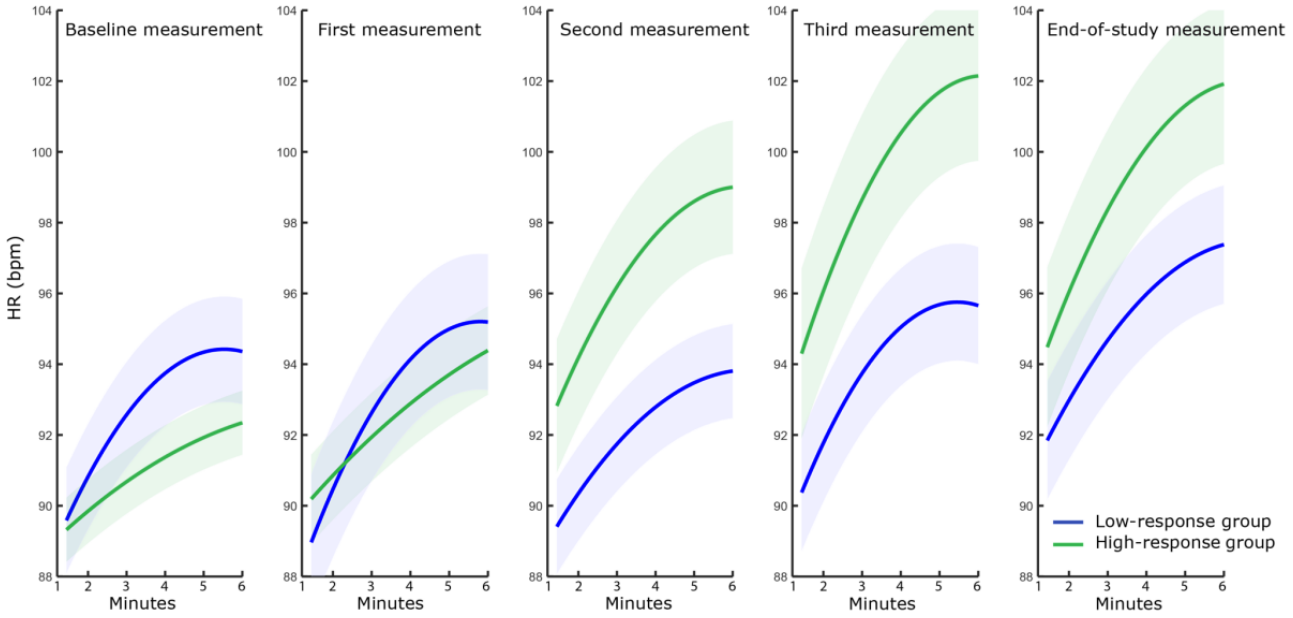
The HR_{rec} during every minute of recuperation increased throughout CR for the high-response group, while no change was observed in the low-response group (HR_{rec1}: *P*<.001, HR_{rec2}: *P*<.001, HR_{rec3}: *P*<.001, HR_{rec4}: *P*<.001, and HR_{rec5}: *P*=.02). This evolution in HR_{rec} throughout CR resulted in a higher HR_{rec} in the high-response group at the end of the study (HR_{rec2}: *P*=.02, HR_{rec3}: *P*=.02, HR_{rec4}: *P*=.03, and HR_{rec5}: *P*=.02). In other words, patients from the high-response group will recuperate faster at the end of CR.

Digital Cardiac Biomarkers to Capture the Dynamic Behavior of Heart Rate

The dynamic behavior of heart rate during a standardized exercise and subsequent recovery phase were investigated to further understand the difference in cardiac response between the groups. Hereto, four different models were fitted to the heart rate data. The quadratic polynomial fit obtained the best goodness of fit and was extracted from the heart rate data; the resulting coefficients, a_{poly} and b_{poly}, were analyzed. Both coefficients determine the shape and steepness of the curve, thereby characterizing the speed of heart rate increase during the 6MWT and, thus, the response of the heart to exercise. These innovative digital cardiac biomarkers were studied for differences between the groups during every session, as well as for the difference in their progression throughout CR.

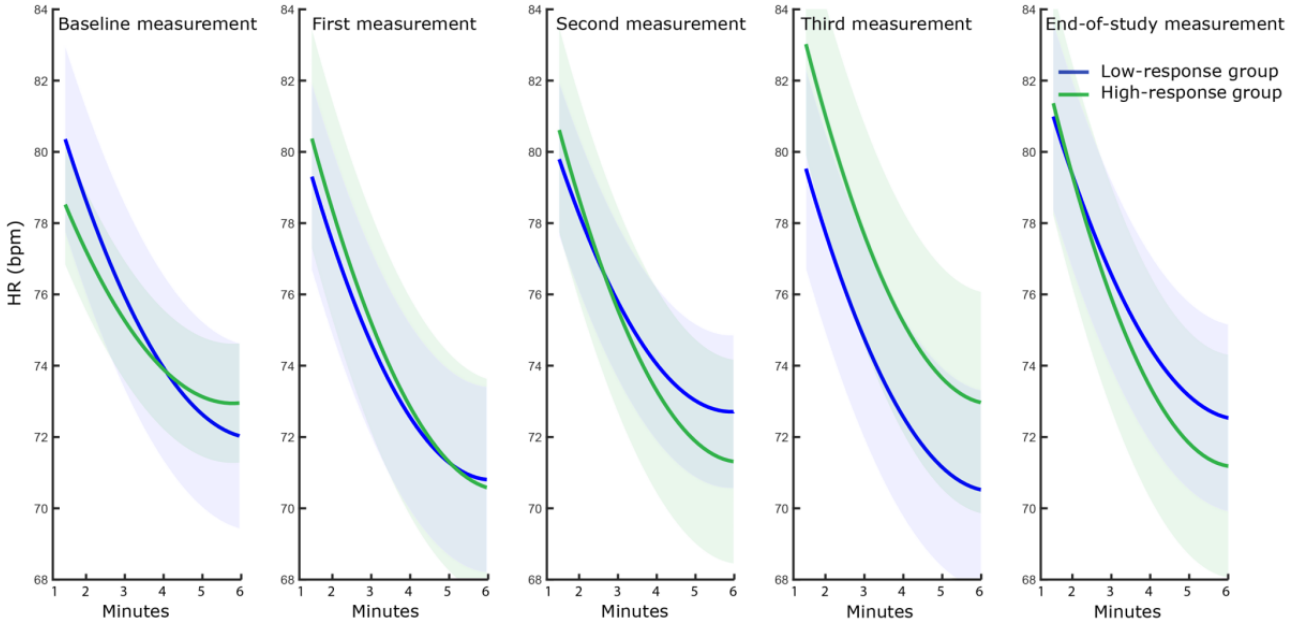
Figure 5 shows the mean quadratic polynomial fit ($f(x) = a_{poly}x^2 + b_{poly}x + c$) to the heart rate behavior during all five 6MWTs performed throughout the CR program. The behavior of heart rate measured during the walking phase evolved differently throughout CR in both groups. This is reflected in the difference in evolution for the polynomial coefficients, a_{poly} and b_{poly}, between the groups ($F_{4,96}$=5.691, *P*=.008, partial η^2=.133 and $F_{4,160}$= 4.302, *P*=.01, partial η^2=.175). The higher values of a_{poly} and b_{poly} during baseline and the subsequent 6MWT in the low-response group indicate that the heart rate of these patients increases faster at the start of CR (-.00013 vs -.000027 and .065 vs .022, *P*<.001). However, toward the end of CR, the heart rate of the high-response group shows an increase and eventually catches up with the low-response group, showing a similar heart rate behavior during walking.

Figure 5. Mean quadratic polynomial fit to the changes in heart rate (HR) during all sessions while walking. The line represents the mean fit and the shadows represent the SD. bpm: beats per minute.



The behavior of heart rate during the subsequent recovery phase was analyzed for both groups (see Figure 6). The coefficient, b_{poly}, extracted from the polynomial fit, showed a different evolution throughout CR between the groups ($F_{4,236}$=3.367, P=.01, partial η^2=.054). More specifically, the behavior of heart rate during recovery changed over time for the high-response group, while the low-response group showed no change in heart rate behavior during recovery (P<.001).

Figure 6. Mean quadratic polynomial fit to the changes in heart rate (HR) during all sessions while recuperating. The line represents the mean fit and the shadows represent the SD. bpm: beats per minute.



Discussion

Principal Findings

The findings of this observational study indicate that cardiac response to exercise in patients following a CR program plays a role in the level of their response to training in terms of distance walked. To our knowledge, this is the first study to describe the longitudinal follow-up of a CR patient population using wearable sensor technology during a repeated, standardized, submaximal activity test. Our 3-month follow-up period allows for novel in-depth insights into the cardiac response at rest, during exercise, and during recovery. Cardiac response is one of the possible confounders affecting the response to CR in patients. Therefore, investigating this cardiac response in a typical CR program aids in understanding the mechanisms behind different response rates. The wearable sensor technology enabled continuous monitoring of heart rate to derive both traditional heart rate parameters, commonly used in practice, and innovatively derived parameters that can function as digital cardiac biomarkers. The descriptive analysis of this longitudinally collected dataset investigated the difference

in cardiac response between two patient populations following CR, who are respectively characterized by low and high improvement in functional capacity during CR.

Evolution of Commonly Used Static Heart Rate Parameters Throughout Cardiac Rehabilitation

Previous research showed that the response to training depends on the cardiac output or chronotropic response, as this determines the increase in muscle blood flow during exercise [17,28]. Cardiac patients often depend on increasing their heart rate during exercise, as an increase in stroke volume is often limited due to the impaired cardiac output [29]. Heart rate parameters measured before, during, and after exercise contain information on how the heart responds to exercise. First, the effect of CR on the evolution of HR_{rest}, HR_{peak}, and HR_{rec} in patients showing a high or low response to exercise intervention was analyzed. These heart rate parameters, which are commonly used in practice, comprise the heart rate–related information from the resting, walking, and recuperation phases, respectively, into one single value.

HR_{peak} is a cardiac biomarker that evolves differently throughout CR for both response groups. The high-response group showed an increase in HR_{peak}, while the low-response group showed no significant change. At a first glance, these results might appear to be opposite of the results from previous research [30-32], which showed that an increase in functional capacity is accompanied by a decrease in HR_{peak}. However, they studied the heart rate response during a maximal exercise test, while in this study patients performed a submaximal test. The increase in HR_{peak} that we observed is accompanied by an increased distance. Hence, it is difficult to conclude whether the increased HR_{peak} is a result of the CR program or the fact that the walking distance also increases. Previous research showed that exercise intolerance causes a limitation in cardiac response in cardiac patients, as they achieved a higher heart rate compared to controls during submaximal exercise at a similar workload [33,34]. Effort is a measure of the walking intensity during the 6MWT. Both groups showed a similar effort. Although there was an increase in effort during the 6MWT throughout the CR, only the *response* group was characterized by an increase in HR_{peak}. This indicated that the cardiac response in the *nonresponse* group was limited, leading to exercise intolerance throughout the CR program. Additionally, a correction was made for the distance. After correcting HR_{peak} for distance, a significant decrease in $HR_{peak\text{-}dist}$ was seen for both groups. Thus, the results indicate that a cautious interpretation of heart rate parameters is necessary when studying the effects of CR on cardiac response during a submaximal exercise test.

The difference between the two groups in heart rate recovery, as captured by HR_{rec}, was investigated. According to Qiu et al, HR_{rec} measured after the 6MWT is considered to be a powerful prognostic indicator in cardiovascular disease [35]. Cardiovascular patients are characterized by a slower HR_{rec} due to an attenuated autonomous nervous system. This decrease in HR_{rec} can be partially restored, as following a CR program improves the impaired recovery [36]. The results showed that HR_{rec} changed differently throughout the CR program for each group. In addition, our study showed no differences between the absolute HR_{rec} values during the first four sessions, but indicated a difference in HR_{rec} between the groups during the last session. The faster recuperation toward the end of CR could indicate the positive effect on the attenuated vagal reactivation within the high-response group, while this effect is absent in the low-response group.

To summarize, these heart rate parameters show that the effect of CR on the different response groups is also reflected in differences in cardiac response. The cardiac system of the high-response group adapts better to exercise throughout CR compared to the low-response group.

Evolution of Digital Cardiac Biomarkers Throughout Cardiac Rehabilitation

To further understand the difference in cardiac response, the dynamic behavior of heart rate during the 6MWT and subsequent recuperation phase was also investigated. This innovative type of heart rate analysis allows in-depth insights by reflecting changes within shorter time spans during a standardized activity test. The heart rate response is representative of the ability of the autonomic nervous system to meet the hemodynamic demands during exercise. The heart rate acceleration at the onset of exercise is often modelled by an exponential curve [37]. Previous research showed that biphasic and sigmoidal curves are suitable to model heart rate behavior due to the increase in sympathetic activity following vagal withdrawal [38,39]. The choice to model the heart rate with exponential and polynomial curves was based on both literature and the heart rate behavior during the 6MWT, which was characterized by a steep increase followed by a steady-state phase as seen in this CR population.

The heart rate behavior during the 6MWT and subsequent recuperation phase evolved differently between the groups throughout CR. The increase in heart rate during the walking phase was steeper in the low-response group at baseline, indicating that in this phase of the CR the autonomous nervous system of these patients is characterized by a superior response to exercise in comparison to the high-response group. In the course of the CR program, the increase in heart rate steepened for the high-response group, eventually catching up with the low-response group. These results are similar to the findings of Schmid et al and Jorde et al, who indicated that the heart rate slope was blunted in subjects with an impaired cardiac response [17,40]. The ability of the *response* group to increase the heart rate slope, and the fact that the heart rate slope remained unchanged in the *nonresponse* group, indicated that an impaired cardiac response could lead to an impaired response to exercise training. Schmid et al showed that nonresponders to exercise training showed poor heart rate recovery, indicative of a disturbed cardiac autonomic status [17]. Our results confirm these findings, as both groups also showed a different evolution in heart rate behavior during the recuperation phase. The high-response group patients recuperated faster toward the end of CR, while the patients in the low-response group showed no change in recuperation rate.

To summarize, this study shows novel differences between groups, as the evolution in heart rate changes differently throughout the CR. Continuous measurements using wearable sensor technology enables the collection of traditional, commonly used heart rate parameters, but also of digital cardiac biomarkers representative of heart rate behavior during and after activity. The latter parameters were derived using a polynomial curve fitting technique, which is an innovative approach to capture heart rate evolution. Our research showed that these digital cardiac biomarkers can differentiate between the low- and high-response groups at baseline; hence, they, together with the traditional heart rate–related parameters, are valuable tools to use in short-term follow-up. Moreover, the results contribute to the development toward a more patient-tailored treatment strategy. Future research should focus on the role of these heart rate–related parameters in predicting outcome. The ability to improve and complement short-term follow-up by using these innovative techniques could make it possible to adjust treatment strategy in time and optimize outcome.

Limitations

This study is an observational study analyzing the characteristics of a typical CR population; as patients were not randomized into different groups, the results should be interpreted as hypothesis generating. A low number of patients received both a baseline and end-of-study CPET measurement. These missing values do not influence the outcome of the study, as a submaximal exercise test is used to determine the progression in functional capacity throughout CR. There are some limitations to performing a median split to divide a patient population into two groups. However, data in this observational study was investigated in the search for trends upon which to base future randomized research. Another point of discussion is that the 6MWT is an effort-dependent test and a greater increase in HR_{peak} could be a consequence of higher effort. Therefore, effort derived from the data collected by the triaxial accelerometer was used to determine whether the high-response group was characterized by a higher effort in comparison to the low-response group. Hills et al state that acceleration is proportional to the net external force involved in an activity and, therefore, more directly reflects the energy cost associated with movement [41]. Moreover, effort is a parameter that is often used in VO_2-max estimations during submaximal exercise [42,43]. Therefore, if the high-response group would have reached a higher percentage of their maximal exercise capacity during the 6MWT, this would have been reflected in a higher effort obtained during the 6MWT; however, this was not the case.

Conclusions

Following CR is, without any doubt, beneficial for cardiovascular patients. However, some patients benefit more from CR as they show a larger improvement in functional capacity in comparison to other patients. This study shows the following:

1. Continuous measurements using wearable sensor technology during standardized activity give novel insights into cardiac response between different response groups.
2. Patients showing a larger increase in functional capacity are characterized by a better improvement in cardiac response. This is in contrast to patients showing a low response to exercise intervention.
3. Innovative analysis approaches allowed us to study the difference in heart rate behavior between the response groups in more detail, showing differences in cardiac response at baseline.
4. The results from this study can be used in future research to investigate whether the outcome of CR can be predicted in order to adjust treatment strategy. Moreover, it is a first step toward the development of a more patient-tailored CR program.

Acknowledgments

This report is part of the Limburg Clinical Research Program (LCRP), Hasselt University (UHasselt)-Ziekenhuis Oost-Limburg (ZOL)-Jessa, supported by the foundation Limburg Sterk Merk (LSM), province of Limburg, Flemish government, UHasselt, ZOL, and Jessa Hospital. We would like to thank the engineers from Holst Centre, imec the Netherlands, for their technical support. We thank the physiotherapists from the cardiac rehabilitation center, ZOL, for their support and guidance during the study. HDC is supported by a doctoral fellowship by the Research Foundation, Flanders (FWO), Belgium (grant number: 1S53616N). SVH is supported by the Flemish Government under the Onderzoeksprogramma Artificiële Intelligentie Vlaanderen program. SVH and CV's research is supported by Agentschap Innoveren en Ondernemen (VLAIO) (150466-OSA+) and imec funds 2017.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Flowchart containing the intervention measures, dropout reasons, and number of patients. 6MWT: 6-minute walking test.
[PNG File , 325 KB - jmir_v22i5e17326_app1.png]

References

1. Ponikowski P, Voors A, Anker S, Bueno H, Cleland J, Coats A. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J 2016:2129-2200. [doi: 10.3410/f.718489795.793497182]
2. Thomas R, Balady G, Banka G, Beckie T, Chiu J, Gokak S, et al. 2018 ACC/AHA Clinical performance and quality measures for cardiac rehabilitation: A report of the American College of Cardiology/American Heart Association Task Force on performance measures. Circ Cardiovasc Qual Outcomes 2018 Apr;11(4):e000037. [doi: 10.1161/hcq.0000000000000037]
3. Yancy C, Jessup M, Bozkurt B, Butler J, Casey DJ, Drazner M, American College of Cardiology Foundation, American Heart Association Task Force on Practice Guidelines. 2013 ACCF/AHA guideline for the management of heart failure: A report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 2013 Oct 15;62(16):e147-e239 [FREE Full text] [doi: 10.1016/j.jacc.2013.05.019] [Medline: 23747642]
4. Turk-Adawi K, Sarrafzadegan N, Grace SL. Global availability of cardiac rehabilitation. Nat Rev Cardiol 2014 Oct;11(10):586-596 [FREE Full text] [doi: 10.1038/nrcardio.2014.98] [Medline: 25027487]
5. Balady GJ, Arena R, Sietsema K, Myers J, Coke L, Fletcher GF, et al. Clinician's guide to cardiopulmonary exercise testing in adults. Circulation 2010 Jul 13;122(2):191-225. [doi: 10.1161/cir.0b013e3181e52e69]
6. Myers J, Prakash M, Froelicher V, Do D, Partington S, Atwood JE. Exercise capacity and mortality among men referred for exercise testing. N Engl J Med 2002 Mar 14;346(11):793-801. [doi: 10.1056/nejmoa011858]
7. Lawler PR, Filion KB, Eisenberg MJ. Efficacy of exercise-based cardiac rehabilitation post-myocardial infarction: A systematic review and meta-analysis of randomized controlled trials. Am Heart J 2011 Oct;162(4):571-584.e2. [doi: 10.1016/j.ahj.2011.07.017] [Medline: 21982647]
8. Taylor RS, Walker S, Smart NA, Piepoli MF, Warren FC, Ciani O, ExTraMATCH II Collaboration. Impact of exercise-based cardiac rehabilitation in patients with heart failure (ExTraMATCH II) on mortality and hospitalisation: An individual patient data meta-analysis of randomised trials. Eur J Heart Fail 2018 Dec;20(12):1735-1743 [FREE Full text] [doi: 10.1002/ejhf.1311] [Medline: 30255969]
9. Taylor RS, Walker S, Smart NA, Piepoli MF, Warren FC, Ciani O, ExTraMATCH II Collaboration. Impact of exercise rehabilitation on exercise capacity and quality-of-life in heart failure: Individual participant meta-analysis. J Am Coll Cardiol 2019 Apr 02;73(12):1430-1443 [FREE Full text] [doi: 10.1016/j.jacc.2018.12.072] [Medline: 30922474]
10. Anderson L, Sharp GA, Norton RJ, Dalal H, Dean S, Jolly K, et al. Home-based versus centre-based cardiac rehabilitation. Cochrane Database Syst Rev 2017 Jun 30;6:CD007130 [FREE Full text] [doi: 10.1002/14651858.CD007130.pub4] [Medline: 28665511]
11. Keteyian SJ, Brawner CA, Savage PD, Ehrman JK, Schairer J, Divine G, et al. Peak aerobic capacity predicts prognosis in patients with coronary heart disease. Am Heart J 2008 Aug;156(2):292-300. [doi: 10.1016/j.ahj.2008.03.017] [Medline: 18657659]
12. Kavanagh T, Mertens DJ, Hamm LF, Beyene J, Kennedy J, Corey P, et al. Prediction of long-term prognosis in 12,169 men referred for cardiac rehabilitation. Circulation 2002 Aug 06;106(6):666-671. [doi: 10.1161/01.cir.0000024413.15949.ed] [Medline: 12163425]
13. Kavanagh T, Mertens DJ, Hamm LF, Beyene J, Kennedy J, Corey P, et al. Peak oxygen intake and cardiac mortality in women referred for cardiac rehabilitation. J Am Coll Cardiol 2003 Dec 17;42(12):2139-2143. [doi: 10.1016/j.jacc.2003.07.028] [Medline: 14680741]
14. Vanhees L, Fagard R, Thijs L, Amery A. Prognostic value of training-induced change in peak exercise capacity in patients with myocardial infarcts and patients with coronary bypass surgery. Am J Cardiol 1995 Nov;76(14):1014-1019. [doi: 10.1016/s0002-9149(99)80287-2]
15. Aguiar Rosa S, Abreu A, Marques Soares R, Rio P, Filipe C, Rodrigues I, et al. Cardiac rehabilitation after acute coronary syndrome: Do all patients derive the same benefit? Rev Port Cardiol 2017 Mar;36(3):169-176 [FREE Full text] [doi: 10.1016/j.repc.2016.09.011] [Medline: 28238477]
16. Vilela EM, Ladeiras-Lopes R, Ruivo C, Torres S, Braga J, Fonseca M, et al. Different outcomes of a cardiac rehabilitation programme in functional parameters among myocardial infarction survivors according to ejection fraction. Neth Heart J 2019 Jul;27(7-8):347-353 [FREE Full text] [doi: 10.1007/s12471-019-1269-7] [Medline: 30977040]
17. Schmid J, Zurek M, Saner H. Chronotropic incompetence predicts impaired response to exercise training in heart failure patients with sinus rhythm. Eur J Prev Cardiol 2013 Aug;20(4):585-592. [doi: 10.1177/2047487312444368] [Medline: 22492866]
18. Gevaert AB, Adams V, Bahls M, Bowen TS, Cornelissen V, Dörr M, et al. Towards a personalised approach in exercise-based cardiovascular rehabilitation: How can translational research help? A 'call to action' from the Section on Secondary Prevention and Cardiac Rehabilitation of the European Association of Preventive Cardiology. Eur J Prev Cardiol 2019 Oct 04:2047487319877716. [doi: 10.1177/2047487319877716] [Medline: 31581819]
19. ATS Committee on Proficiency Standards for Clinical Pulmonary Function Laboratories. ATS statement: Guidelines for the six-minute walk test. Am J Respir Crit Care Med 2002 Jul 01;166(1):111-117. [doi: 10.1164/ajrccm.166.1.at1102] [Medline: 12091180]

20. Van Helleputte N, Konijnenburg M, Pettine J, Jee D, Kim H, Morgado A, et al. A 345 µW multi-sensor biomedical SoC with bio-impedance, 3-channel ECG, motion artifact reduction, and integrated DSP. IEEE J Solid-State Circuits 2015 Jan;50(1):230-244. [doi: 10.1109/jssc.2014.2359962]
21. Barold SS. Willem Einthoven and the birth of clinical electrocardiography a hundred years ago. Card Electrophysiol Rev 2003 Jan;7(1):99-104. [doi: 10.1023/a:1023667812925] [Medline: 12766530]
22. Varon C, Testelmans D, Buyse B, Suykens J, Van Huffel S. Robust artefact detection in long-term ECG recordings based on autocorrelation function similarity and percentile analysis. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2012 Presented at: Annual International Conference of the IEEE Engineering in Medicine and Biology Society; August 28-September 1, 2012; San Diego, CA p. 3151-3154. [doi: 10.1109/embc.2012.6346633]
23. Moeyersons J, Amoni M, Van Huffel S, Willems R, Varon C. R-DECO: An open-source Matlab based graphical user interface for the detection and correction of R-peaks. PeerJ Comput Sci 2019;5:e226. [doi: 10.7717/peerj-cs.226]
24. Prahm KP, Witting N, Vissing J. Decreased variability of the 6-minute walk test by heart rate correction in patients with neuromuscular disease. PLoS One 2014;9(12):e114273 [FREE Full text] [doi: 10.1371/journal.pone.0114273] [Medline: 25479403]
25. Bai J, Di C, Xiao L, Evenson KR, LaCroix AZ, Crainiceanu CM, et al. An activity index for raw accelerometry data and its comparison with other activity metrics. PLoS One 2016;11(8):e0160644 [FREE Full text] [doi: 10.1371/journal.pone.0160644] [Medline: 27513333]
26. Colley RC, Garriguet D, Janssen I, Craig CL, Clarke J, Tremblay MS. Physical activity of Canadian adults: Accelerometer results from the 2007 to 2009 Canadian Health Measures Survey. Health Rep 2011 Mar;22(1):7-14 [FREE Full text] [Medline: 21510585]
27. Migueles JH, Cadenas-Sanchez C, Ekelund U, Delisle Nyström C, Mora-Gonzalez J, Löf M, et al. Accelerometer data collection and processing criteria to assess physical activity and other outcomes: A systematic review and practical considerations. Sports Med 2017 Sep;47(9):1821-1845 [FREE Full text] [doi: 10.1007/s40279-017-0716-0] [Medline: 28303543]
28. Sullivan MJ, Knight JD, Higginbotham MB, Cobb FR. Relation between central and peripheral hemodynamics during exercise in patients with chronic heart failure. Muscle blood flow is reduced with maintenance of arterial perfusion pressure. Circulation 1989 Oct;80(4):769-781. [doi: 10.1161/01.cir.80.4.769] [Medline: 2791242]
29. Wilson JR, Groves J, Rayos G. Circulatory status and response to cardiac rehabilitation in patients with heart failure. Circulation 1996 Oct 01;94(7):1567-1572. [doi: 10.1161/01.cir.94.7.1567] [Medline: 8840845]
30. Casillas J, Gudjoncik A, Gremeaux V, Aulagne J, Besson D, Laroche D. Assessment tools for personalizing training intensity during cardiac rehabilitation: Literature review and practical proposals. Ann Phys Rehabil Med 2017 Jan;60(1):43-49 [FREE Full text] [doi: 10.1016/j.rehab.2016.01.011] [Medline: 26996956]
31. Diller G, Dimopoulos K, Okonko D, Uebing A, Broberg CS, Babu-Narayan S, et al. Heart rate response during exercise predicts survival in adults with congenital heart disease. J Am Coll Cardiol 2006 Sep 19;48(6):1250-1256 [FREE Full text] [doi: 10.1016/j.jacc.2006.05.051] [Medline: 16979014]
32. Wilmore JH, Stanforth PR, Gagnon J, Rice T, Mandel S, Leon AS, et al. Heart rate and blood pressure changes with endurance training: The HERITAGE Family Study. Med Sci Sports Exerc 2001 Jan;33(1):107-116. [doi: 10.1097/00005768-200101000-00017] [Medline: 11194095]
33. Witte KK, Cleland JG, Clark AL. Chronic heart failure, chronotropic incompetence, and the effects of beta blockade. Heart 2006 Apr;92(4):481-486 [FREE Full text] [doi: 10.1136/hrt.2004.058073] [Medline: 16159968]
34. Zweerink A, van der Lingen AC, Handoko ML, van Rossum AC, Allaart CP. Chronotropic incompetence in chronic heart failure. Circ Heart Fail 2018 Aug;11(8):e004969. [doi: 10.1161/CIRCHEARTFAILURE.118.004969] [Medline: 30354566]
35. Qiu S, Cai X, Sun Z, Li L, Zuegel M, Steinacker JM, et al. Heart rate recovery and risk of cardiovascular events and all‐cause mortality: A meta‐analysis of prospective cohort studies. J Am Heart Assoc 2017 May 09;6(5):e005505 [FREE Full text] [doi: 10.1161/JAHA.117.005505] [Medline: 28487388]
36. Peçanha T, Silva-Júnior ND, de Moraes Forjaz CL. Heart rate recovery: Autonomic determinants, methods of assessment and association with mortality and cardiovascular diseases. Clin Physiol Funct Imaging 2014 Sep;34(5):327-339. [doi: 10.1111/cpf.12102] [Medline: 24237859]
37. Feroldi P, Belleri M, Ferretti G, Veicsteinas A. Heart rate overshoot at the beginning of muscle exercise. Eur J Appl Physiol Occup Physiol 1992;65(1):8-12. [doi: 10.1007/BF01466267] [Medline: 1505545]
38. Krzemiński K, Nazar K, Cybulski G, Niewiadomski W. Endurance training slows down the kinetics of heart rate increase in the transition from moderate to heavier submaximal exercise intensities. Eur J Appl Physiol Occup Physiol 1991;62(4):297-300. [doi: 10.1007/BF00571556] [Medline: 2044541]
39. Thomson RL, Rogers DK, Howe PR, Buckley JD. Effect of acute exercise-induced fatigue on maximal rate of heart rate increase during submaximal cycling. Res Sports Med 2016;24(1):1-15. [doi: 10.1080/15438627.2015.1076414] [Medline: 26289798]

40. Jorde UP, Vittorio TJ, Kasper ME, Arezzi E, Colombo PC, Goldsmith RL, et al. Chronotropic incompetence, beta-blockers, and functional capacity in advanced congestive heart failure: Time to pace? Eur J Heart Fail 2008 Jan;10(1):96-101 [FREE Full text] [doi: 10.1016/j.ejheart.2007.11.006] [Medline: 18096432]
41. Hills AP, Mokhtar N, Byrne NM. Assessment of physical activity and energy expenditure: An overview of objective measures. Front Nutr 2014;1:5 [FREE Full text] [doi: 10.3389/fnut.2014.00005] [Medline: 25988109]
42. De Brabandere A, Op De Beéck T, Schütte KH, Meert W, Vanwanseele B, Davis J. Data fusion of body-worn accelerometers and heart rate to predict VO2max during submaximal running. PLoS One 2018;13(6):e0199509 [FREE Full text] [doi: 10.1371/journal.pone.0199509] [Medline: 29958282]
43. Tönis TM, Gorter K, Vollenbroek-Hutten MM, Hermens H. Comparing VO2max determined by using the relation between heart rate and accelerometry with submaximal estimated VO2max. J Sports Med Phys Fitness 2012 Aug;52(4):337-343. [Medline: 22828454]

Abbreviations

6MWT: 6-minute walking test
ANOVA: analysis of variance
CPET: cardiopulmonary exercise test
CR: cardiac rehabilitation
df_{error}: degrees of freedom for the error term
df_{main}: degrees of freedom for the simple main effect
ECG: electrocardiogram
FWO: Research Foundation, Flanders
HR_{peak}: peak heart rate
$HR_{peak-dist}$: peak heart rate corrected for the distance walked
HR_{rec}: recovery heart rate
HR_{rest}: resting heart rate
LCRP: Limburg Clinical Research Program
LSM: Limburg Sterk Merk
MUSEIC: MUlti SEnsor IC
UHasselt: Hasselt University
VLAIO: Agentschap Innoveren en Ondernemen
VO_2: peak oxygen uptake
ZOL: Ziekenhuis Oost-Limburg

Edited by G Eysenbach; submitted 07.12.19; peer-reviewed by IV George, H Kemps, K Ng; comments to author 15.01.20; revised version received 06.03.20; accepted 10.04.20; published 20.05.20.

Please cite as:
De Cannière H, Smeets CJP, Schoutteten M, Varon C, Van Hoof C, Van Huffel S, Groenendaal W, Vandervoort P
Using Biosensors and Digital Biomarkers to Assess Response to Cardiac Rehabilitation: Observational Study
J Med Internet Res 2020;22(5):e17326
URL: http://www.jmir.org/2020/5/e17326/
doi:10.2196/17326
PMID:32432552

Original Paper

A Brief Music App to Address Pain in the Emergency Department: Prospective Study

Peter R Chai[1,2,3,4], MA, MD; Emily Schwartz[5], BA; Mohammad Adrian Hasdianda[6], MD; Desiree R Azizoddin[2], PsyD; Anna Kikut[6], BA; Guruprasad D Jambaulikar[6], MPH, MD; Robert R Edwards[5], PhD; Edward W Boyer[1,4], MD, PhD; Kristin L Schreiber[5], MD, PhD

[1]Division of Medical Toxicology, Department of Emergency Medicine, Brigham and Women's Hospital, Boston, MA, United States

[2]Department of Psychosocial Oncology and Palliative Care, Dana Farber Cancer Institute, Boston, MA, United States

[3]The Koch Institute for Integrated Cancer Research, Massachusetts Institute of Technology, Boston, MA, United States

[4]The Fenway Institute, Boston, MA, United States

[5]Department of Anesthesiology, Perioperative, and Pain Medicine, Brigham and Women's Hospital, Boston, MA, United States

[6]Department of Emergency Medicine, Brigham and Women's Hospital, Boston, MA, United States

Corresponding Author:
Peter R Chai, MA, MD
Division of Medical Toxicology
Department of Emergency Medicine
Brigham and Women's Hospital
75 Francis St
Boston, MA, 02115
United States
Phone: 1 617 732 5640
Email: pchai@bwh.harvard.edu

Abstract

Background: Emergency physicians face the challenge of relieving acute pain daily. While opioids are a potent treatment for pain, the opioid epidemic has ignited a search for nonopioid analgesic alternatives that may decrease the dose or duration of opioid exposure. While behavioral therapies and complementary medicine are effective, they are difficult to deploy in the emergency department. Music is a potential adjunctive therapy that has demonstrated effectiveness in managing pain.

Objective: Our objective was to understand the feasibility and potential for an effect of a novel music app to address acute pain and anxiety in patients admitted to an emergency department observation unit.

Methods: This prospective cohort study enrolled patients admitted to an emergency department observation unit with pain who had received orders for opioids. We gathered baseline pain and psychosocial characteristics including anxiety, sleep disturbance, and pain catastrophizing using validated questionnaires. Participants received a smartphone-based music intervention and listened to the music in either a supervised (research assistant–delivered music session 3 times during their stay) or unsupervised manner (music used ad lib by participant). The app collected premusic and postmusic pain and anxiety scores, and participants provided qualitative feedback regarding acceptability of operating the music intervention.

Results: We enrolled 81 participants and randomly assigned 38 to an unsupervised and 43 to a supervised group. Mean pain in both groups was 6.1 (1.8) out of a possible score of 10. A total of 43 (53%) reported previous use of music apps at home. We observed an overall modest but significant decrease in pain (mean difference –0.81, 95% CI –0.45 to –1.16) and anxiety (mean difference –0.72, 95% CI –0.33 to –1.12) after music sessions. Reduction of pain and anxiety varied substantially among participants. Individuals with higher baseline pain, catastrophizing (about pain), or anxiety reported greater relief. Changes in pain were correlated to changes in anxiety (Pearson ρ=0.3, *P*=.02) but did not vary between supervised and unsupervised groups. Upon conclusion of the study, 46/62 (74%) reported they liked the music intervention, 57/62 (92%) reported the app was easy to use, and 49/62 (79%) reported they would be willing to use the music intervention at home.

Conclusions: A smartphone-based music intervention decreased pain and anxiety among patients in an emergency department observation unit, with no difference between supervised and unsupervised use. Individuals reporting the greatest reduction in pain after music sessions included those scoring highest on baseline assessment of catastrophic thinking, suggesting there may

be specific patient populations that may benefit more from using music as an analgesic adjunct in the emergency department. Qualitative feedback suggested that this intervention was feasible and acceptable by emergency department patients.

(J Med Internet Res 2020;22(5):e18537) doi:10.2196/18537

KEYWORDS

music therapy; pain; smartphone; technology; telemedicine; emergency service, hospital

Introduction

Background

The number of opioid analgesics prescribed in the United States is now recognized as unsustainably and dangerously high [1,2], with no indication that the experience of pain among individuals has improved as a result. Three problematic consequences of increased opioid prescription are (1) overdose deaths, with over 47,000 individuals dying from opioid-related overdose in 2017 [3] (approximately 128 individuals per day); (2) an epidemic of opioid use disorder, with escalation to injection opioid use and increasing numbers of opioid-related bacterial infections [4]; and (3) an inherent conflict of clinical priorities in the emergency department (ED), where providers confront the competing pressures of reducing exposure to opioid analgesics while still managing acutely painful conditions, with limited alternatives [5,6]. Nonopioid analgesics, increasingly the first-line treatment for certain acutely painful conditions, often provide an inadequate response. Enhancing the response to nonopioid (or opioid) analgesics through adjunctive use of behavior modulatory therapies (eg, cognitive behavioral therapy) [7] or complementary medicine (eg, yoga, acupuncture) is an attractive strategy [8,9], but these therapies are difficult to deploy in the ED. Therefore, exploration of other more feasible and scalable alternative strategies is needed to reduce or replace opioid pharmacotherapy in this context.

One potential adjunctive therapy is listening to music, a nearly universal experience shared across cultures, ethnicities, and races [10]. The ubiquity of music across cultures and known human history may arise from its capacity to improve affect and mood. Functional magnetic resonance imaging studies have shown that pleasurable music listening is associated with increased dopaminergic neurotransmission in the nucleus accumbens, a common reward pathway in the brain [11,12]. Music also improves affect, with at least one study showing a beneficial effect on both pain and anxiety in the perioperative setting [13].

The favorable neurochemistry induced by listening to music may enable an innovative behavioral intervention. Machine learning has identified features (ie, musical “genes”) associated with increased dopaminergic neurotransmission and improved affect while listening to pleasurable music [14,15]. These features intrinsic to music itself have formed the basis for commercial music selection algorithms such as the music genome project [16]. Similar techniques have been applied in the development of novel music programs to create new signatures of pleasurable music to be played from a smartphone app [17,18].

Objective

The objective of this study was to evaluate the feasibility and effectiveness of a smartphone-based music intervention for acute pain in patients presenting to the ED. We specifically sought to measure changes in pain and anxiety associated with 10-minute music listening sessions delivered in either a supervised or an ad hoc fashion to patients with pain symptoms in an ED observation unit.

Methods

Participant Recruitment

In this prospective study, we recruited patients who were admitted to an ED observation unit in a quaternary, urban, level I trauma academic ED. Study staff identified potential participants by screening the electronic medical record and liaising with ED clinical staff. English-speaking patients with ED observation unit stays of up to 48 hours and who had received orders for opioid analgesia were eligible to participate. We excluded patients who had a pacemaker, had hearing loss, were on contact precautions, had previously enrolled in the study, were unstably housed, or had a significant active medical or psychiatric illness. A trained research assistant (RA) explained the study in full and gauged potential participants’ interest in the study. Interested participants provided verbal consent. The Partners Human Research Committee (Institutional Review Board of Partners HealthCare, Boston, MA, USA) approved this protocol.

Participants provided basic demographic information and underwent baseline pain and psychosocial assessments using validated measures (see section on measures) using a study phone with direct connection to a REDCap data capture system version 9.5 (REDCap Consortium) [19]. Next, participants were introduced to and given instructions for operating the novel music app on a study smartphone (iPhone 6; Apple Inc, Cupertino, CA, USA). We were interested in 2 modalities of use of the music app. First, we wanted to understand the use of music by participants if they were not given formal guidance as to how often and how long they should use music for. Second, we wanted to determine the feasibility of using a regimented prescription of music supervised by an RA. We therefore randomly assigned participants via a predetermined, computer-generated random list into unsupervised versus supervised intervention use groups. Participants randomly assigned to the unsupervised group received a smartphone with the app preloaded and paired noise-canceling headphones and were encouraged to use the music intervention app on an as-needed basis. Participants also received contact information for RAs so they could seek help to troubleshoot any potential issues with the app or the devices used. For participants

randomly assigned to the supervised intervention group, a study RA reminded and assisted the participant to use the music app up to 3 times during their ED observation unit stay (approximately every 4 hours).

Measures

We used several measures to assess baseline patient characteristics.

Pain

The Brief Pain Inventory (BPI) is a 9-item, self-report, validated measure that assesses patients' pain severity (least, worst, and average) and functional interference in the preceding week using an 11-point Likert scale with higher scores indicating greater pain or functional impact, or both [20].

Psychosocial Characteristics

We used the Patient-Reported Outcomes Measurement Information System (PROMIS) depression (8 items), sleep disturbance (8 items), and anxiety (7 items) short forms [21]. All PROMIS measures were scored on a 5-point Likert scale with higher scores indicating higher symptomatology. The Pain Catastrophizing Scale (PCS) comprises 13 items exploring individuals' pain catastrophizing divided into 3 subscales for rumination about pain, magnification of pain, and perceived helplessness to do anything about pain [22]. The PCS uses a 5-point Likert scale with cumulative scores ranging from 0 to 52. The Perceived Stress Scale consists of 10 items to measure subjective stress and uses a 5-point Likert scale for each item, where higher total scores indicate greater stress (range 0-40) [23]. We measured somatization using the 7-item somatization subscale from the Brief Symptom Inventory, where response options were scored on a 5-point Likert scale, with higher scores indicating greater somatization [24].

The Music Intervention

Participants interacted with a smartphone-based music app that used 5 separate commissioned instrumental tracks intended to be relaxing (Unwind; Bose Corporation, Boston, MA, USA) [25]. These music tracks were composed using a backbone of musical genes derived through a machine learning algorithm that selected key musical features expected to induce subjective "thrills" among individuals, combined with human-composed music. Users were presented with a preview screen that allowed them to sample 5 different tracks and select their favorite track. The same 5 tracks were presented to each participant. Next, participants used noise-canceling headphones to listen to 10 minutes of the selected track. Participants rated their *pain* and *anxiety* using a 10-point Likert scale at the beginning and end of the 10-minute music session. At completion of the study procedures, participants were guided through an RA-administered survey to assess the feasibility of operating the music app and formative reflections on using music in the ED observation unit. We collected data on the total opioid administered during the participants' ED observation unit stay from the electronic medical record and converted does to milligram morphine equivalents (MMEs). We then divided the MME amount by the number of hours the patient spent in the ED observation unit to calculate a normalized opioid utilization score (MME/h) for each participant during the study.

Data Analysis

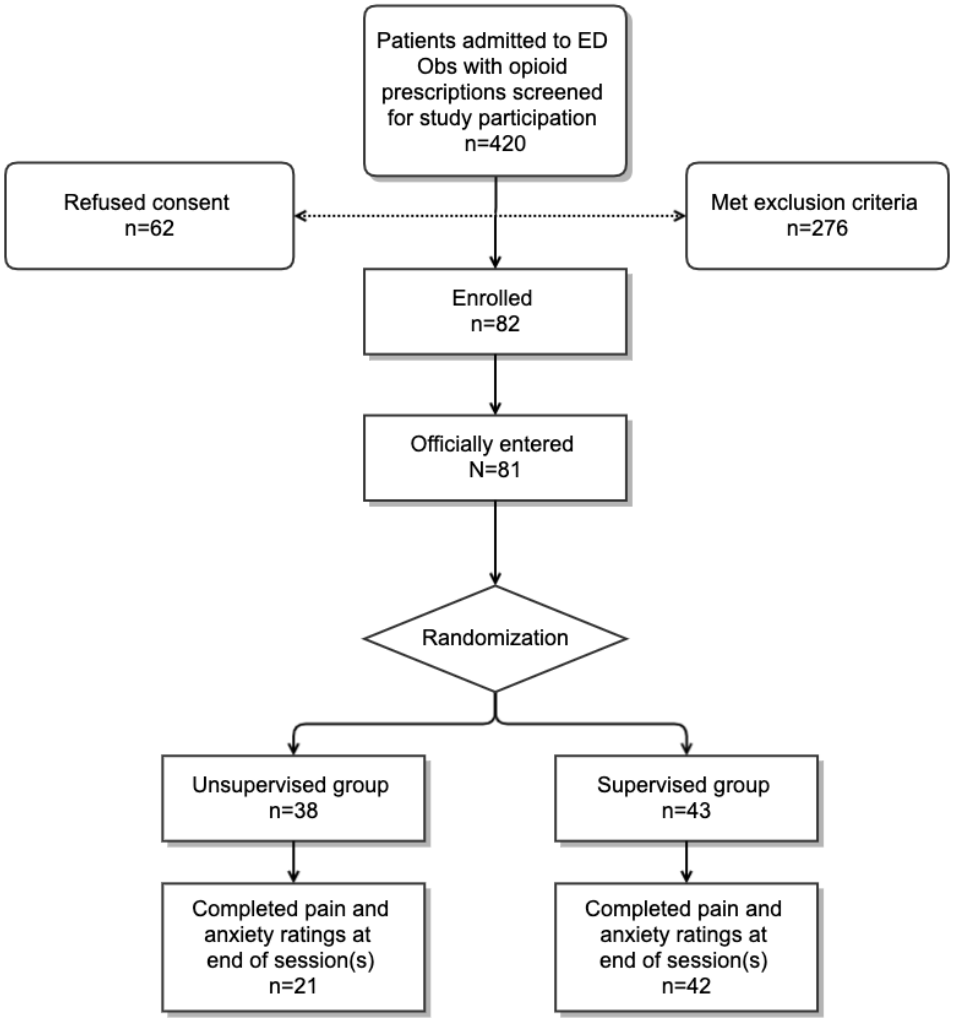
We summarized participant characteristics using frequencies and percentages for categorical variables, and mean and standard deviation or median and interquartile range for continuous variables, according to normality of distribution. For patient-reported scores of baseline and postmusic session pain and anxiety, we calculated mean values for each patient. We compared mean premusic versus mean postmusic scores using a nonparametric paired test (related-samples Wilcoxon signed rank test). We averaged the change in pain scores by subtracting mean end pain score from mean beginning pain score; we assessed this value (change in pain with music) for correlation with baseline patient characteristics, including psychosocial (anxiety, depression, catastrophizing, stress) and general pain severity and interference (BPI) questionnaires, using Spearman or Pearson correlations, as appropriate. All statistical tests were 2-tailed, and level of significance was set at Cronbach α=.05. We performed all analyses using IBM SPSS 25 (IBM Corporation).

Results

Participant Characteristics

We screened 420 individuals, of whom 144 were eligible for participation (Figure 1). Of these, we enrolled 82 patients. Common reasons for nonparticipation were lack of interest in research, other concomitant ongoing clinical assessments, and the inability to operate a smartphone.

Figure 1. Study flow diagram. ED Obs: emergency department observation unit.



The mean age of participants in the total sample (N=81) was 43.84 (SD 15) years, with a larger proportion of men (57, 70%) than women (24, 30%; Table 1). Reported pain in the previous week was high (BPI mean pain score 6.12, SD 1.8, out of 10), and at baseline 14 (17%) of participants were taking opioids on a regular basis. Participants reported relatively elevated scores on depression, anxiety, stress, and somatization (Table 1). Similarly, pain catastrophizing scores were higher than in previous reports of healthy volunteers [22] and similar to ranges reported among patients with chronic pain. We identified no difference between the supervised and unsupervised groups in terms of baseline pain or psychosocial characteristics (P>.05).

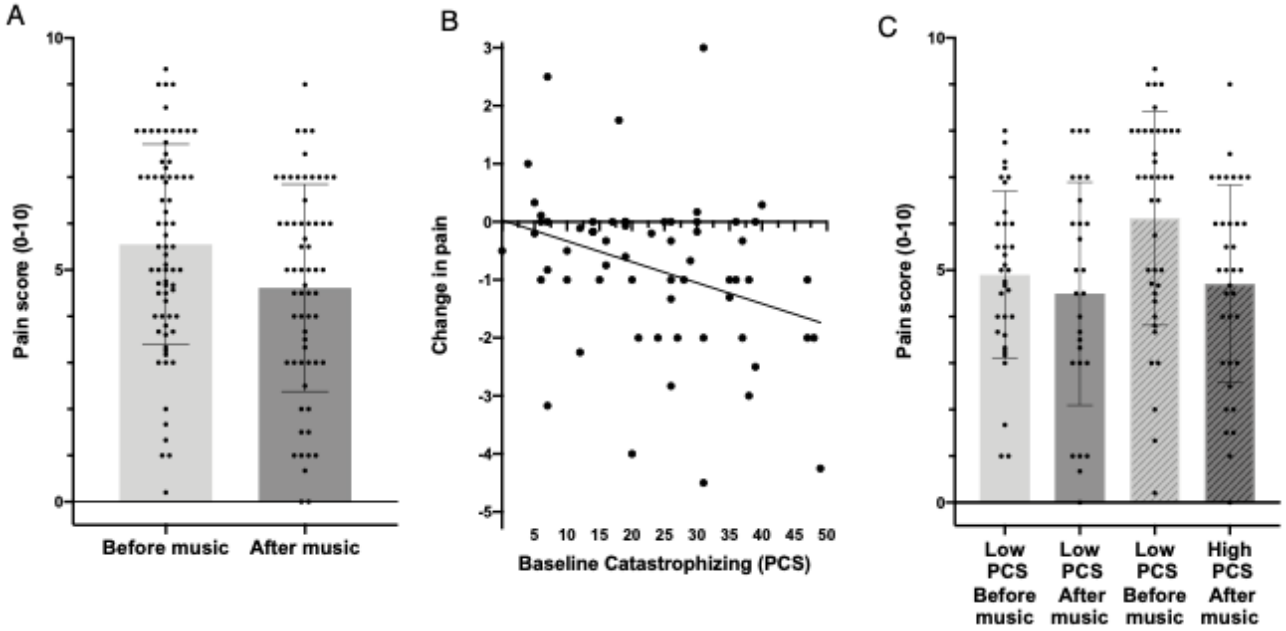
Table 1. Patient sample characteristics (N=81).

Characteristics	Values
Age (years), mean (SD)	43.84 (15)
Female sex, n (%)	24 (30)
Taking opioids at baseline, n (%)	14 (17)
Baseline Brief Pain Inventory (BPI) score, mean (SD)	
BPI current	6.17 (2.1)
BPI worst	8.61 (2.2)
BPI least	3.02 (2.4)
BPI mean	6.12 (1.8)
BPI interference	6.22 (2.4)
Patient-Reported Outcomes Measurement Information System score, mean (SD)	
Anxiety (score range 7-35)	19.13 (6.2)
Depression (score range 8-40)	17.38 (7.5)
Sleep Disturbance (score range 8-40)	29.22 (6.4)
Somatization (Brief Symptom Inventory; score range 0-35), mean (SD)	12.61 (4.0)
Perceived Stress Scale (score range 0-40)	18.83 (3.5)
Pain Catastrophizing Scale (score range 0-52), mean (SD)	
Rumination	8.35 (4.7)
Magnification	3.28 (2.5)
Helplessness	10.48 (6.4)
Total	22.11 (12.5)

Change in Pain During a Music Session

A total of 73 participants successfully attempted the music sessions, with a mean of 3 music sessions and a range of 1 to 9 music sessions per participant. We obtained fewer postmusic pain ratings than premusic ratings, with the missed postmusic ratings frequently being due to the participant falling asleep while listening to the music, as noted by RAs in supervised administrations. Comparison of mean pain scores before and after music interventions revealed a significant overall reduction in pain scores in most individuals (Figure 2, part A; related-samples Wilcoxon signed rank test, $P<.001$). The mean change in pain scores with music was modest (mean difference –0.81, 95% CI –0.45 to –1.16), with considerable variability among participants. To explore participant characteristics associated with the greatest benefit, we correlated the magnitude change in pain scores to baseline characteristics. Those with higher current pain reported on the BPI experienced greater pain relief with the use of music (Spearman $\rho=-0.37$, $P=.004$).

Figure 2. Effect of music on pain in emergency department patients. (A) Comparison of mean pain scores before and after music interventions revealed a significant overall reduction in pain scores in most individuals (related-samples Wilcoxon signed rank test, *P*<.001). The mean change in pain score with music was modest (mean difference –0.81, 95% CI –0.45 to –1.16) with considerable variability among participants. (B) Baseline Pain Catastrophizing Scale (PCS) score measured before music interventions was correlated with the amount of change in pain after music session. Higher baseline PCS scores were associated with a larger decrease in pain after music session (Spearman ρ=–0.39, *P*=.009). (C) Change in pain after music among low and high catastrophizers. Patients with high baseline PCS score (>20) had a greater decrease in pain (mean difference –1.2, SD 1.4) after music session than those with lower PCS (mean difference –0.3, SD 1.1; independent-samples *t* test=–2.9, *P*=.005).



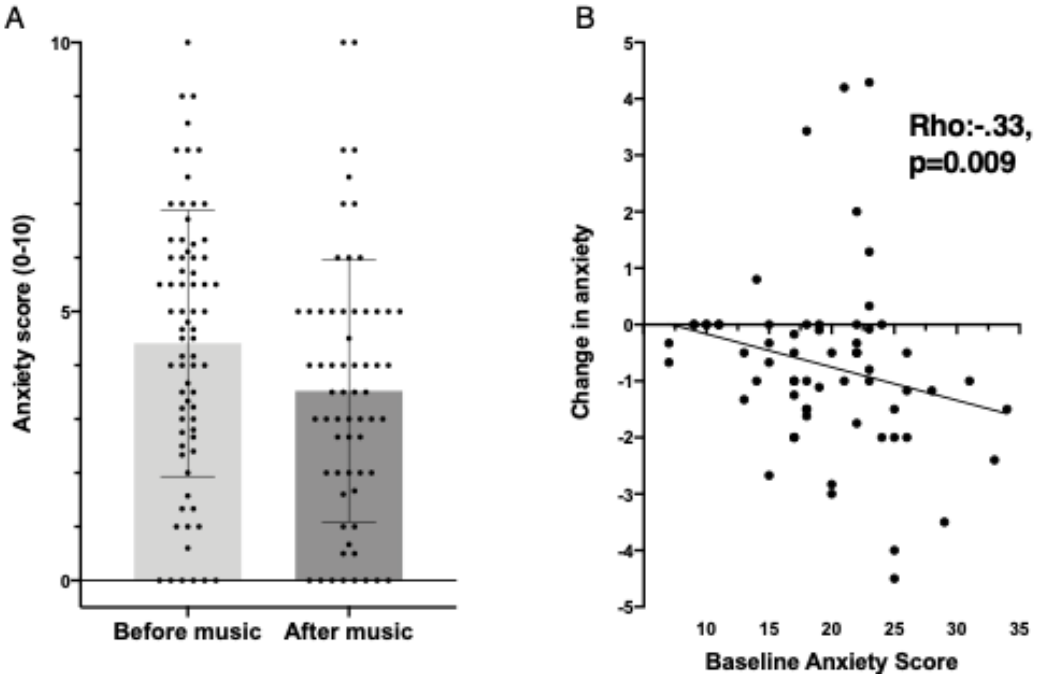
Patient Characteristics Associated With Pain Reduction

The psychosocial variables anxiety, depression, sleep disturbance, and catastrophizing showed significant intercorrelation to each other, as well as a significant association with reported acute pain (Multimedia Appendix 1). The PCS score at baseline was inversely correlated with the change in pain during the music session (difference between the end and beginning pain scores; (Spearman ρ=–0.39, *P*=.009; Figure 2, part B); participants with initial PCS scores greater than 20 were more likely to report a significant decrease in pain with the music intervention (Figure 2, part C). Other psychosocial variables (anxiety, depression, somatization, stress) were not significantly related to the degree of music-induced change in pain. Pain reduction did not differ between the supervised and unsupervised groups (*P*>.05).

Change in Anxiety During a Music Session

Comparison of mean anxiety scores reported before and after the music intervention revealed a significant overall reduction in anxiety in most individuals (Figure 3, part A; related-samples Wilcoxon signed rank test, *P*<.001). As with pain, the mean magnitude of change in anxiety scores with music was modest (mean difference –0.72, 95%CI –0.33 to –1.12), with considerable variability among participants. Greater reduction in anxiety scores significantly correlated with baseline anxiety, with those reporting high baseline anxiety also showing a greater reduction in anxiety with music (Figure 3, part B). Other psychosocial variables (depression, somatization, stress, catastrophizing) were not significantly related to the degree of music-induced change in anxiety. Change in pain was modestly related to change in anxiety (Pearson ρ=0.30, *P*=.02). Anxiety reduction did not differ between the supervised and unsupervised groups (*P*>.05).

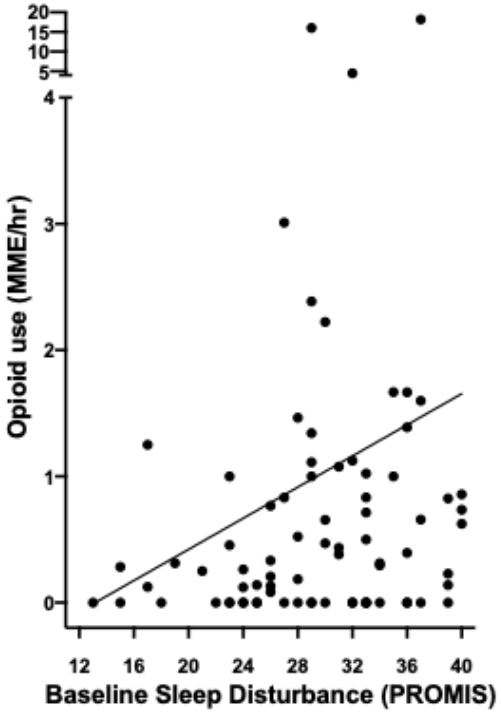
Figure 3. Effect of music on anxiety in emergency department patients. (A) Comparison of mean anxiety scores reported before and after the music listening intervention revealed a significant overall reduction in anxiety among individuals (related-samples Wilcoxon signed rank test, *P*<.001). The mean magnitude change in anxiety score with music was modest (mean difference –0.72, 95% CI –0.33 to –1.12) with considerable variability among participants. (B) Higher baseline anxiety scores were associated with a greater reduction in anxiety scores before and after the music session (Spearman ρ=0.3, *P*=.02).



Relationship of Baseline Characteristics to Opioid Use While in the ED Observation Unit

As a secondary exploratory analysis, we examined whether certain patient characteristics were associated with higher opioid use during the ED observation unit stay. A higher reported baseline sleep disturbance was associated with increased opioid consumption (Figure 4). Other baseline patient characteristics were not significantly related to changes in opioid use.

Figure 4. Variation in opioid use among participants with baseline sleep disturbance. Baseline sleep disturbance (measured by the Patient-Reported Outcomes Measurement Information System [PROMIS] short form) was associated with a higher amount of opioid consumption (milligram morphine equivalent [MME]) per hour while in the emergency department (Spearman ρ=0.24, *P*=.03).



Feasibility and User Experience

Analyses of the exit survey data revealed that most participants had used music as therapy before (43/62, 69%), and many participants had used other music apps in the past (23/62, 37%). The majority of participants reported they liked the music app (46/62, 74%), with the majority (43/62, 69%) saying they used it to relax during their ED stay, and some (15/62, 24%) endorsed using the music to sleep (Table 2). Most participants (49/62, 79%) said they would use it at home.

Table 2. Summary of the exit survey impressions of the Unwind App (n=62).

User experience	Values, n (%)
App was easy to use	57 (92%)
Liked the music app	46 (74%)
Purpose of app use: relax	43 (69%)
Purpose of app use: sleep	15 (24%)
Would use the app at home	49 (79%)

Discussion

Principal Findings

These findings suggest that music may be used to decrease pain and anxiety among individuals admitted to an ED observation unit. Individuals who had increasing catastrophizing at baseline (measured by the PCS) experienced greater decreases in pain and anxiety after listening to the music app. This demonstrates that music interventions may potentially be applied with greater effect in individuals with higher catastrophizing in the ED. Our findings are consistent with other investigations that described many other sources of music to be associated with the reduction of postoperative or procedural pain in individuals with chronic pain [26-32]. Additionally, the feasibility of using this music intervention in both a supervised and an unsupervised model suggests that this intervention can be used as a prescribed, targeted intervention in the ED or used ad hoc in individuals who may experience pain outside of the ED.

Music has several hypothesized effects on pain. First, music may modulate affect by impacting reward pathways, as changes in dopaminergic neurotransmission are associated with pleasurable or relaxing music. Music may decrease other pain-augmenting factors such as anxiety. Second, music may alter the nociceptive processing of pain by changing pain tolerance and threshold [33-35]. Alternatively, music may divert attention from acute pain to a different, pleasurable stimulus [10,18,30-33,35-38]. Regardless of the mechanism, the universality of music—every culture on the globe has some form of musical expression—coupled with the near ubiquity of personal music players, smartphones, and streaming music services makes music an attractive potential adjunctive therapy for pain.

Our data suggest that this app-based, semistructured music intervention can be employed during brief ED observation unit admissions to reduce pain and anxiety in patients with acute pain. Participants reported pain scores in real time and operated a music app in both supervised and unsupervised conditions. Additionally, participants' pain and anxiety scores were not different between the supervised and unsupervised groups. Integrated into emergency care, the use of music may afford an adjunctive treatment to those patients with higher catastrophizing and anxiety, which may be particularly relevant in emergency care settings [5].

While music reduces pain and anxiety in controlled settings [26-32], our investigation demonstrated that these effects may also exist in the midst of the unpredictable environment of the ED. These decreases in pain and anxiety, while modest, may meaningfully alter the ED experience for some individuals by serving as an adjunct to opioid analgesics. While this study was not designed to evaluate the effect of music on opioid consumption, its potential combination with other nonopioid interventions could contribute to decreased opioid use while in the ED and decrease opioid dosage units prescribed at discharge [10].

Notably, the modulation of pain produced by this music intervention was greatest in individuals with high levels of catastrophizing (negative cognitions and rumination about pain). Similarly, reduction of anxiety during the intervention was largest among those with highest baseline anxiety levels. The finding that music has a greater impact on those individuals with high levels of baseline catastrophizing anxiety is notable, as these traits are often associated with greater resistance to pharmacological manipulation, as well as pain persistence. Other previous studies among pain patients have found an association between greater catastrophizing scores and greater efficacy of behavioral interventions directed at affect, mood, and catastrophizing [39-41]. Whether the pain-sparing effect of music among high catastrophizers is due to alteration of sensory perception of nociceptive stimuli, as some laboratory-based investigations have suggested [17,18,28], or whether music modulates affect and thus indirectly decreases the impact of painful stimuli [27,36] remains to be clearly defined. Attentional modulation of (distraction away from) painful stimuli may be an important aspect of how music provides pain relief [41], although it seems unlikely to be the sole mechanism, as distraction alone has not led to pain reductions in similar populations [42].

Our exit survey data of patients at the end of the study support a possible role of affect modulation in music's effect. For example, one participant described that they continued to experience pain while listening to the music but felt "better" about their pain. Others reported decreased anxiety about their pain, while some notably reported that they preferred the use

of music over receiving an opioid. Interestingly, one observation from the study was that many participants fell asleep during the listening session, consistent with many previous studies that have noted an important connection between sleep deprivation and pain [43]. These findings suggest that the interplay between music and pain is complex. We anticipate that future investigations will address both the mechanisms in which music therapy may integrate within the biopsychosocial model of pain, and the implementation science behind operationalizing a recommendation for music to address the experience of pain.

Limitations

This study had several limitations. First, we conducted our study at a single urban academic ED. Experiences regarding complaint and psychosocial profiles of study participants and the resources available in ED observation units may be different in other types of ED settings, thus limiting generalizability to similar cohorts. Second, participants operating the app did not always rate postintervention pain and anxiety each time they used it. Anecdotally, participants reported that listening to the music made them fall asleep—itself a highly beneficial effect—thus leaving them unable to complete the postassessment, and possibly underestimating the degree of pain and anxiety reduction. This, taken together with the observation that individuals with the greatest sleep disturbance used the most opioids, suggests that future work should explore the use of music intervention to augment sleep in individuals with pain. Third, because our primary objective was the effectiveness and feasibility of using music to reduce pain and anxiety, we employed a pre-post assessment over multiple sessions, and there was thus not a control group in this study. Fourth, the use of the music app was not directly tied to opioid administration in the ED observation unit. We did not encourage participants or nursing staff to access the music app at particular times proximal to their scheduled opioid administration. Future investigations should consider offering music in strategic times near opioid administration to understand whether the analgesic effects of music can diminish the times individuals request opioids. This also made it impossible to investigate the impact of music on opioid use during the study period.

Conclusion

A smartphone-based intervention built on machine learning–based musical knowledge produced a significant decrease in both pain and anxiety among ED observation unit patients with acute pain. Participants experienced decreased pain and anxiety regardless of whether they used the music intervention on an as-needed basis for pain or under direct supervision. Individuals with higher catastrophizing scores at baseline gained more benefit from the music intervention. Further randomized, controlled studies evaluating the effect of different types of music intervention on pain modulation, pain catastrophizing, and opioid use are needed [43]. Careful assessment of baseline characteristics may help further delineate those who are most likely to derive analgesic benefit from music. Ultimately, our results demonstrate that a music intervention is a feasible adjunctive therapy for individuals who have pain in the ED, as participants willingly used a music intervention when offered to them.

Acknowledgments

Portions of this paper were presented at the 2019 Hawaii International Conference on System Sciences.

EWB received grant funding from the US National Institute on Drug Abuse (NIDA; R01DA047236); PRC received grant funding from NIDA (K23DA044874) and the Hans and Mavis Lopater Psychosocial Fund, and investigator-initiated research grants from Gilead Sciences, Inc, and E Ink Corporation; and KLS received grant funding from the US National Institute of General Medical Sciences (R35GM128691).

The described music intervention and technical support were provided by Bose, Inc free of charge.

Authors' Contributions

PRC, EWB, RRE, and KLS conceived the study. EWB obtained technical support for the study. PRC, AK, GDJ, and MAH conducted the study. AK, GDJ, and MAH were responsible for data acquisition. KLS, ES, MAH, GDJ, and PRC conducted data analysis. PRC drafted the initial manuscript. KLS, ES, DRA, MAH, and GDJ contributed key edits to the final manuscript. All authors contributed key edits to the final manuscript and approved the contents in the manuscript. PRC takes responsibility for the manuscript as a whole.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Correlation table between psychosocial factors and pain.
[PNG File , 152 KB - jmir_v22i5e18537_app1.png]

References

1. Okie S. A flood of opioids, a rising tide of deaths. N Engl J Med 2010 Nov 18;363(21):1981-1985. [doi: 10.1056/NEJMp1011512] [Medline: 21083382]
2. Bohnert ASB, Valenstein M, Bair MJ, Ganoczy D, McCarthy JF, Ilgen MA, et al. Association between opioid prescribing patterns and opioid overdose-related deaths. JAMA 2011 Apr 06;305(13):1315-1321. [doi: 10.1001/jama.2011.370] [Medline: 21467284]
3. Scholl L, Seth P, Kariisa M, Wilson N, Baldwin G. Drug and opioid-involved overdose deaths - United States, 2013-2017. MMWR Morb Mortal Wkly Rep 2018 Jan 04;67(5152):1419-1427 [FREE Full text] [doi: 10.15585/mmwr.mm675152e1] [Medline: 30605448]
4. Cicero TJ, Ellis MS, Surratt HL, Kurtz SP. The changing face of heroin use in the United States: a retrospective analysis of the past 50 years. JAMA Psychiatry 2014 Jul 01;71(7):821-826. [doi: 10.1001/jamapsychiatry.2014.366] [Medline: 24871348]
5. Motov SM, Khan AN. Problems and barriers of pain management in the emergency department: are we ever going to get better? J Pain Res 2008 Dec 09;2:5-11 [FREE Full text] [Medline: 21197290]
6. Chang AK, Bijur PE, Esses D, Barnaby DP, Baer J. Effect of a single dose of oral opioid and nonopioid analgesics on acute extremity pain in the emergency department: a randomized clinical trial. JAMA 2017 Nov 07;318(17):1661-1667 [FREE Full text] [doi: 10.1001/jama.2017.16190] [Medline: 29114833]
7. Roditi D, Robinson ME. The role of psychological interventions in the management of patients with chronic pain. Psychol Res Behav Manag 2011;4:41-49 [FREE Full text] [doi: 10.2147/PRBM.S15375] [Medline: 22114534]
8. Ward L, Stebbings S, Cherkin D, Baxter GD. Yoga for functional ability, pain and psychosocial outcomes in musculoskeletal conditions: a systematic review and meta-analysis. Musculoskeletal Care 2013 Dec;11(4):203-217. [doi: 10.1002/msc.1042] [Medline: 23300142]
9. Burns JR, Kram JJF, Xiong V, Stark Casadont JM, Mullen TA, Conway N, et al. Utilization of acupuncture services in the emergency department setting: a quality improvement study. J Patient Cent Res Rev 2019;6(2):172-178 [FREE Full text] [doi: 10.17294/2330-0698.1688] [Medline: 31414028]
10. Chai PR, Carreiro S, Ranney ML, Karanam K, Ahtisaari M, Edwards R, et al. Music as an adjunct to opioid-based analgesia. J Med Toxicol 2017 Sep;13(3):249-254 [FREE Full text] [doi: 10.1007/s13181-017-0621-9] [Medline: 28646359]
11. Blum K, Chen TJH, Chen ALH, Madigan M, Downs BW, Waite RL, et al. Do dopaminergic gene polymorphisms affect mesolimbic reward activation of music listening response? Therapeutic impact on Reward Deficiency Syndrome (RDS). Med Hypotheses 2010 Mar;74(3):513-520. [doi: 10.1016/j.mehy.2009.10.008] [Medline: 19914781]
12. Blum K, Simpatico T, Febo M, Rodriquez C, Dushaj K, Li M, et al. Hypothesizing music intervention enhances brain functional connectivity involving dopaminergic recruitment: common neuro-correlates to abusable drugs. Mol Neurobiol 2017 Jul;54(5):3753-3758 [FREE Full text] [doi: 10.1007/s12035-016-9934-y] [Medline: 27246565]
13. Chlan LL, Weinert CR, Heiderscheit A, Tracy MF, Skaar DJ, Guttormson JL, et al. Effects of patient-directed music intervention on anxiety and sedative exposure in critically ill patients receiving mechanical ventilatory support: a randomized clinical trial. JAMA 2013 Jun 12;309(22):2335-2344 [FREE Full text] [doi: 10.1001/jama.2013.5670] [Medline: 23689789]
14. Dannenberg RB, Hu N. Pattern discovery techniques for music audio. J New Music Res 2003 Jun 1;32(2):153-163. [doi: 10.1076/jnmr.32.2.153.16738]
15. Morris JW, Powers D. Control, curation and musical experience in streaming music services. Creative Ind J 2015 Oct 20;8(2):106-122. [doi: 10.1080/17510694.2015.1090222]
16. Glaser WT, Westergren TB, Stearns JP, Kraft JM, inventors; Pandora Media, Inc, assignee. Consumer item matching method and system. US patent 7,003,515 B1. 2006 Feb 21. URL: https://docs.google.com/viewer?url=patentimages.storage.googleapis.com/pdfs/US7003515.pdf [accessed 2020-05-04]
17. Salimpoor VN, Benovoy M, Larcher K, Dagher A, Zatorre RJ. Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. Nat Neurosci 2011 Feb;14(2):257-262. [doi: 10.1038/nn.2726] [Medline: 21217764]
18. Menon V, Levitin DJ. The rewards of music listening: response and physiological connectivity of the mesolimbic system. Neuroimage 2005 Oct 15;28(1):175-184. [doi: 10.1016/j.neuroimage.2005.05.053] [Medline: 16023376]
19. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
20. Tan G, Jensen MP, Thornby JI, Shanti BF. Validation of the Brief Pain Inventory for chronic nonmalignant pain. J Pain 2004 Mar;5(2):133-137. [doi: 10.1016/j.jpain.2003.12.005] [Medline: 15042521]
21. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. Med Care 2007 May;45(5 Suppl 1):S3-S11 [FREE Full text] [doi: 10.1097/01.mlr.0000258615.42478.55] [Medline: 17443116]
22. Sullivan MJL, Bishop SR, Pivik J. The Pain Catastrophizing Scale: development and validation. Psychol Assess 1995;7(4):524-532. [doi: 10.1037/1040-3590.7.4.524]
23. Roberti J, Harrington L, Storch E. Further psychometric support for the 10-item version of the Perceived Stress Scale. J Coll Couns 2006;9(2):135-147. [doi: 10.1002/j.2161-1882.2006.tb00100.x]

24. Derogatis LR, Melisaratos N. The Brief Symptom Inventory: an introductory report. Psychol Med 2009 Jul 09;13(3):595-605. [doi: 10.1017/s0033291700048017]
25. Chai PR, Schreiber KL, Taylor SW, Jambaulikar GD, Kikut A, Hasdianda MA, et al. The feasibility and acceptability of a smartphone-based music intervention for acute pain. Proc Annu Hawaii Int Conf Syst Sci 2019 Jan 08;2019:3917-3925 [FREE Full text] [Medline: 30700971]
26. De Silva AP, Niriella MA, Nandamuni Y, Nanayakkara SD, Perera KRP, Kodisinghe SK, et al. Effect of audio and visual distraction on patients undergoing colonoscopy: a randomized controlled study. Endosc Int Open 2016 Nov;4(11):E1211-E1214 [FREE Full text] [doi: 10.1055/s-0042-117630] [Medline: 27853748]
27. Liu Y, Petrini MA. Effects of music therapy on pain, anxiety, and vital signs in patients after thoracic surgery. Complement Ther Med 2015 Oct;23(5):714-718. [doi: 10.1016/j.ctim.2015.08.002] [Medline: 26365452]
28. Özer N, Karaman Özlü Z, Arslan S, Günes N. Effect of music on postoperative pain and physiologic parameters of patients after open heart surgery. Pain Manag Nurs 2013 Mar;14(1):20-28. [doi: 10.1016/j.pmn.2010.05.002] [Medline: 23452523]
29. Drzymalski DM, Tsen LC, Palanisamy A, Zhou J, Huang C, Kodali BS. A randomized controlled trial of music use during epidural catheter placement on laboring parturient anxiety, pain, and satisfaction. Anesth Analg 2017 Feb;124(2):542-547. [doi: 10.1213/ANE.0000000000001656] [Medline: 27984250]
30. Yeo JK, Cho DY, Oh MM, Park SS, Park MG. Listening to music during cystoscopy decreases anxiety, pain, and dissatisfaction in patients: a pilot randomized controlled trial. J Endourol 2013 Apr;27(4):459-462 [FREE Full text] [doi: 10.1089/end.2012.0222] [Medline: 23009573]
31. Onieva-Zafra MD, Castro-Sánchez AM, Matarán-Peñarrocha GA, Moreno-Lorenzo C. Effect of music as nursing intervention for people diagnosed with fibromyalgia. Pain Manag Nurs 2013 Jun;14(2):e39-e46. [doi: 10.1016/j.pmn.2010.09.004] [Medline: 23108015]
32. Gutgsell KJ, Schluchter M, Margevicius S, DeGolia PA, McLaughlin B, Harris M, et al. Music therapy reduces pain in palliative care patients: a randomized controlled trial. J Pain Symptom Manage 2013 May;45(5):822-831 [FREE Full text] [doi: 10.1016/j.jpainsymman.2012.05.008] [Medline: 23017609]
33. Garza-Villarreal EA, Jiang Z, Vuust P, Alcauter S, Vase L, Pasaye EH, et al. Music reduces pain and increases resting state fMRI BOLD signal amplitude in the left angular gyrus in fibromyalgia patients. Front Psychol 2015;6:1051 [FREE Full text] [doi: 10.3389/fpsyg.2015.01051] [Medline: 26257695]
34. Dobek CE, Beynon ME, Bosma RL, Stroman PW. Music modulation of pain perception and pain-related activity in the brain, brain stem, and spinal cord: a functional magnetic resonance imaging study. J Pain 2014 Oct;15(10):1057-1068. [doi: 10.1016/j.jpain.2014.07.006] [Medline: 25077425]
35. Guétin S, Giniès P, Siou DKA, Picot M, Pommié C, Guldner E, et al. The effects of music intervention in the management of chronic pain: a single-blind, randomized, controlled trial. Clin J Pain 2012 May;28(4):329-337. [doi: 10.1097/AJP.0b013e31822be973] [Medline: 22001666]
36. Roy M, Peretz I, Rainville P. Emotional valence contributes to music-induced analgesia. Pain 2008 Jan;134(1-2):140-147. [doi: 10.1016/j.pain.2007.04.003] [Medline: 17532141]
37. McCaffrey R, Freeman E. Effect of music on chronic osteoarthritis pain in older people. J Adv Nurs 2003 Dec;44(5):517-524. [doi: 10.1046/j.0309-2402.2003.02835.x] [Medline: 14651700]
38. Verhoeven K, Van Damme S, Eccleston C, Van Ryckeghem DML, Legrain V, Crombez G. Distraction from pain and executive functioning: an experimental investigation of the role of inhibition, task switching and working memory. Eur J Pain 2011 Sep;15(8):866-873. [doi: 10.1016/j.ejpain.2011.01.009] [Medline: 21397536]
39. Litt MD, Shafer DM, Ibanez CR, Kreutzer DL, Tawfik-Yonkers Z. Momentary pain and coping in temporomandibular disorder pain: exploring mechanisms of cognitive behavioral treatment for chronic pain. Pain 2009 Sep;145(1-2):160-168 [FREE Full text] [doi: 10.1016/j.pain.2009.06.003] [Medline: 19553018]
40. Sturgeon JA. Psychological therapies for the management of chronic pain. Psychol Res Behav Manag 2014;7:115-124 [FREE Full text] [doi: 10.2147/PRBM.S44762] [Medline: 24748826]
41. Schreiber KL, Martel MO, Shnol H, Shaffer JR, Greco C, Viray N, et al. Persistent pain in postmastectomy patients: comparison of psychophysical, medical, surgical, and psychosocial characteristics between patients with and without pain. Pain 2013 May;154(5):660-668 [FREE Full text] [doi: 10.1016/j.pain.2012.11.015] [Medline: 23290256]
42. Campbell CM, Witmer K, Simango M, Carteret A, Loggia ML, Campbell JN, et al. Catastrophizing delays the analgesic effect of distraction. Pain 2010 May;149(2):202-207 [FREE Full text] [doi: 10.1016/j.pain.2009.11.012] [Medline: 20188470]
43. Haack M, Simpson N, Sethna N, Kaur S, Mullington J. Sleep deficiency and chronic pain: potential underlying mechanisms and clinical implications. Neuropsychopharmacology 2020 Jan;45(1):205-216. [doi: 10.1038/s41386-019-0439-z] [Medline: 31207606]

Abbreviations

BPI: Brief Pain Inventory
ED: emergency department
MME: milligram morphine equivalent

PCS: Pain Catastrophizing Scale
PROMIS: Patient-Reported Outcomes Measurement Information System
RA: research assistant

Edited by G Eysenbach; submitted 05.03.20; peer-reviewed by S Carreiro, R Lee, X Garcia-Eroles; comments to author 24.03.20; revised version received 18.04.20; accepted 30.04.20; published 20.05.20.

Please cite as:
Chai PR, Schwartz E, Hasdianda MA, Azizoddin DR, Kikut A, Jambaulikar GD, Edwards RR, Boyer EW, Schreiber KL
A Brief Music App to Address Pain in the Emergency Department: Prospective Study
J Med Internet Res 2020;22(5):e18537
URL: http://www.jmir.org/2020/5/e18537/
doi:10.2196/18537
PMID:32432550

Original Paper

Detection of Atrial Fibrillation Using a Ring-Type Wearable Device (CardioTracker) and Deep Learning Analysis of Photoplethysmography Signals: Prospective Observational Proof-of-Concept Study

Soonil Kwon[1*], MD; Joonki Hong[2*], MSc; Eue-Keun Choi[1], MD, PhD; Byunghwan Lee[3], MSc; Changhyun Baik[3], MSc; Euijae Lee[1], MD; Eui-Rim Jeong[4], PhD; Bon-Kwon Koo[1], MD, PhD; Seil Oh[1], MD, PhD, FHRS, FESC; Yung Yi[2], PhD

[1]Seoul National University Hospital, Seoul, Republic of Korea

[2]Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea

[3]Sky Labs Inc, Seongnam, Republic of Korea

[4]Department of Information and Communication Engineering, Hanbat National University, Daejeon, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Eue-Keun Choi, MD, PhD
Seoul National University Hospital
101 Daehak-ro, Jongno-gu
Seoul
Republic of Korea
Phone: 82 2 2072 0688
Email: choiek417@gmail.com

Abstract

Background: Continuous photoplethysmography (PPG) monitoring with a wearable device may aid the early detection of atrial fibrillation (AF).

Objective: We aimed to evaluate the diagnostic performance of a ring-type wearable device (CardioTracker, CART), which can detect AF using deep learning analysis of PPG signals.

Methods: Patients with persistent AF who underwent cardioversion were recruited prospectively. We recorded PPG signals at the finger with CART and a conventional pulse oximeter before and after cardioversion over a period of 15 min (each instrument). Cardiologists validated the PPG rhythms with simultaneous single-lead electrocardiography. The PPG data were transmitted to a smartphone wirelessly and analyzed with a deep learning algorithm. We also validated the deep learning algorithm in 20 healthy subjects with sinus rhythm (SR).

Results: In 100 study participants, CART generated a total of 13,038 30-s PPG samples (5850 for SR and 7188 for AF). Using the deep learning algorithm, the diagnostic accuracy, sensitivity, specificity, positive-predictive value, and negative-predictive value were 96.9%, 99.0%, 94.3%, 95.6%, and 98.7%, respectively. Although the diagnostic accuracy decreased with shorter sample lengths, the accuracy was maintained at 94.7% with 10-s measurements. For SR, the specificity decreased with higher variability of peak-to-peak intervals. However, for AF, CART maintained consistent sensitivity regardless of variability. Pulse rates had a lower impact on sensitivity than on specificity. The performance of CART was comparable to that of the conventional device when using a proper threshold. External validation showed that 94.99% (16,529/17,400) of the PPG samples from the control group were correctly identified with SR.

Conclusions: A ring-type wearable device with deep learning analysis of PPG signals could accurately diagnose AF without relying on electrocardiography. With this device, continuous monitoring for AF may be promising in high-risk populations.

Trial Registration: ClinicalTrials.gov NCT04023188; https://clinicaltrials.gov/ct2/show/NCT04023188

(J Med Internet Res 2020;22(5):e16443) doi:10.2196/16443

KEYWORDS

atrial fibrillation; deep learning; diagnosis; photoplethysmography; wearable electronic devices

Introduction

Atrial fibrillation (AF) is the most common cardiac arrhythmia, and its prevalence has rapidly increased, especially in the elderly population [1]. In view of this trend, 17.9 million adults are expected to develop AF in Europe by the year 2060 [2]. The socioeconomic burdens of AF are also increasing rapidly in line with its prevalence such that annual medical expenses associated with AF have risen at least five fold in the last decade [3]. Considering the serious complications of AF, early diagnosis and proper management are important.

However, the early detection of AF is challenging owing to its paroxysmal nature [4]. This characteristic makes single electrocardiography (ECG) screening no better than pulse palpation to detect silent AF [5]. Besides, the diagnosis of early AF tends to be delayed because the condition is often asymptomatic [6]. This unmet need for early AF detection may be relieved by continuously monitoring the cardiac rhythm in high-risk populations [7]. However, the cost and invasiveness of an implantable loop recorder limit its use as a continuous monitoring device.

Recently, photoplethysmography (PPG) has been assessed to generate a novel biosignal to monitor AF [8-11]. Compared with ECG, PPG has advantages in terms of accessibility and applicability to wearable or mobile devices [12]. In contrast to an inconvenient traditional strategy to confirm cardiac arrhythmia whereby the patient needs to visit the hospital and undergo ECG, PPG can be easily performed at home using a smartphone or a wearable device. Recently, the Apple Heart Study showed that this strategy has the potential to detect underlying AF in the general population [13]. Moreover, AF detection with PPG can be accurate with deep learning analysis, without relying on ECG [14]. However, the accuracy of PPG depends on the site of measurement [15]. The finger has the highest amplitude and the smallest pulse peak time and reflection index for PPG compared with other body parts; thus, it provides maximum information that can be analyzed [16]. The finger also receives a more abundant supply of arterial blood than the wrist and is easier to affix sensors, hence improving signal quality. As a result, collection of PPG signals from the finger is likely to yield better signal quality than that from the wrist. Therefore, a ring-type wearable device may be a more suitable candidate for the acquisition of PPG signals than a wrist-type wearable device, such as the Apple Watch. Thus, we hypothesized that a ring-type wearable device monitoring PPG data would have high diagnostic performance in the detection of AF. This study aimed to develop a ring-type wearable device (CardioTracker, CART) to detect AF with deep learning analysis of PPG signals and to evaluate its diagnostic performance in patients with AF.

Methods

Study Design and Population

This was a prospective, single-center, observational cohort study conducted from 2018 to 2019. The flowchart of this study is illustrated in Multimedia Appendix 1. Adult patients (aged ≥20 years) with persistent AF who were admitted for elective direct-current cardioversion were eligible for the study. The patients were excluded from the study if their cardiac rhythm just before the cardioversion was not AF. The participants who met the eligibility criteria were introduced to the study and enrolled after obtaining informed consent. The recruitment process was consecutive, and measurements were performed in the order of consent to participate in the study.

For direct-current cardioversion, electric shocks of 100-200 J with a biphasic defibrillator were delivered by paddles under light sedation. The cardiac rhythms before and after the cardioversion were validated with 12-lead ECG read by three cardiologists. If there was a discrepancy, a senior electrophysiologist (EKC or EL) assessed the final cardiac rhythm. Both before and after the cardioversion, each participant was at rest in the supine position and PPG and simultaneous single-lead ECG were recorded over 15 min. We did not measure PPG signals after the shock delivery for those in whom cardioversion was unsuccessful. The study protocol was approved by the Institutional Review Board of Seoul National University Hospital and adhered to the Declaration of Helsinki (approval no: 1801-081-916). The study has been registered at ClinicalTrials.gov (NCT04023188).

Measurements

This study used PPG measurements by CART (Sky Labs Inc, Seongnam, Republic of Korea) and a conventional medical-grade pulse oximeter (iDAQ-400 with PPG-AMP and P400, PhysioLab Inc, Busan, Republic of Korea) as the two index tests and synchronized single-lead ECG (lead I) as the reference standard. The two devices recorded PPG signals simultaneously (CART at the proximal phalanx and the conventional device at the fingertip). The participant chose a finger that was the most comfortable for PPG measurements and wore CART. Wearing CART was not very different from wearing a conventional ring. However, to ensure proper signal quality, other fingers were selected if there were scars, thick skin, or tremors. Moreover, three different sized CART devices were prepared for proper contact between the skin and PPG sensors. The measurements of both PPG signals and the single-lead ECG were synchronized. The PPG signals from the conventional device and the single-lead ECG were tracked and recorded by monitoring equipment kept at the bedside, whereas the PPG signals from CART were wirelessly transmitted to a research-purpose smartphone in real time (Figure 1). The rhythms of PPG data were confirmed and labeled by reading synchronized single-lead ECG strips. The cardiac rhythms were classified into sinus rhythm (SR) or AF. Here, SR included a case where there existed premature atrial or ventricular beats.

During the measurement, the participant was required to lie still in bed to minimize motion artifacts.

We applied a bandpass filter (0.2-18 Hz) to the PPG signals, recorded them at a sampling frequency of 50 Hz, and exported them in XML format for preprocessing. Examples of PPG data from CART and the conventional pulse oximeter are illustrated in Multimedia Appendix 2.

For data augmentation [14,17], each 15-min PPG datum was divided into 30-s samples with 20-s overlaps. For the deep learning process, every sample was labelled as AF or SR according to the rhythm of its synchronized single-lead ECG.

Figure 1. Demonstration of photoplethysmography (PPG) monitoring by CardioTracker (CART). CART measures PPG signals at the proximal phalanx and wirelessly transmits the data to the linked smartphone, which can monitor the PPG signals in real-time, and the deep learning algorithm suggests a possible diagnosis.

Ring-Type Wearable Device

CART has been developed to collect and analyze PPG signals from the proximal phalanx. The measurement is based on the reflective method [18], using high-intensity green light-emitting diodes (LEDs) and photodiodes (PDs) embedded inside. It analyzes the PPG signals using a deep learning algorithm with a convolutional neural network (CNN) [17,19], which involves nine neural layers, where the top two are fully connected. The neural network was trained by the Adam optimizer [20]. Dropout and L2 regularization were performed to prevent overfitting [21]. The softmax outputs of the neural network were gently calibrated with temperature scaling to exhibit the diagnostic confidence of SR and AF for every testing PPG sample [22].

To compare the CNN to nondeep learning algorithms, CART also uses a linear-kernel support vector machine (SVM) [23], as it has been shown to have the best diagnostic performance among nondeep learning algorithms [24]. For the features of SVM, we used root mean square of the successive differences of RR intervals with Shannon entropy (RMSSD+ShE) [25], autocorrelation [9], and the ensemble of the previous two (RMSSD+ShE and autocorrelation).

Multimedia Appendix 3 illustrates CART and its wireless charging station. The LEDs and PDs of CART were designed to be located under the finger. CART is made of surgical steel and is waterproof and dustproof (IP58 grade). It collects PPG signals at the proximal phalanx by measuring the reflected lights under the finger. It can monitor PPG signals over 60 hours continuously and can store data for up to 10 hours. It can upload the data via Bluetooth to a mobile or cloud server and can be charged wirelessly with its cradle.

Diagnostic Performance Analysis

We evaluated the baseline characteristics of the study population, including demographics, comorbidities, use of antiarrhythmic agents and anticoagulants, CHA_2DS_2-VASc

scores, and history of AF. Multiple five-fold cross-validation processes were used to perform the training and testing processes of the deep learning algorithm. Each validation process randomly assigned 80% of the total participants for training and the other 20% for testing. The validation process was repeated 10 times for each combination of training and testing datasets, resulting in a total of 50 validation processes for evaluating overall diagnostic performance. For the deep learning process, entire 30-s PPG samples were used without pre-extracted features in training or testing. For a given PPG sample, the deep learning algorithm identified SR or AF, whichever had higher diagnostic confidence. There were no indeterminate or missing data in the two index tests and the reference standard.

To evaluate the diagnostic performance of CART according to the length of the PPG samples, we generated 25-, 20-, 15-, 10-, and 5-s PPG samples from the raw data and repeated the analysis. To investigate whether its performance is affected by the variability of peak-to-peak intervals or the pulse rate of PPG signals, we calculated both the coefficient of variation of the peak-to-peak intervals and the pulse rate for every sample. The diagnostic performance according to the variability and pulse rate was evaluated. As the use of a ring-type device for PPG measurement is not fully understood yet, we compared the two index tests (CART and conventional device) by performing the same analysis but with different PPG measurements. We also assessed the accuracy of the diagnostic performance by trying different subsets of PPG samples; a PPG sample was allowed to be tested by the deep learning algorithm only if its diagnostic confidence was higher than a certain threshold level. For external validation of CART with the deep learning algorithm, 20 healthy subjects with SR were additionally recruited. The PPG signals were measured by CART and processed according to the same protocol in each subject. Considering the random characteristics of deep learning, we repeated the testing 10 times.

Statistical Analysis

The Kolmogorov-Smirnov test was used to check for normal distribution of clinical variables. The data are presented as mean (SD) for age, median with IQR for body mass index and CHA_2DS_2-VASc score, or n (%) for other variables. We obtained cross-tabulation from the validation process and calculated the sensitivity, specificity, positive-predictive value (PPV), negative-predictive value (NPV), and diagnostic accuracy (the ratio between the number of correct cases and the total number of tests). A receiver operating characteristic curve was constructed, and the area under the curve (AUC) with 95% CI was calculated using the diagnostic confidence (CNN) or features (SVM). The mean pulse rate was compared between SR and AF samples using the Student *t* test. All statistical analyses were two-sided, and $P<.05$ was considered statistically significant. The data were analyzed using SPSS version 22.0 (IBM Corp, Armonk, New York, USA).

Results

Baseline Characteristics

The baseline characteristics of the study population are illustrated in Table 1. A total of 100 participants (81 male participants, 81%; mean age 63.8 years, SD 8.5; median CHA_2DS_2-VASc score 2) were enrolled in this study. We collected a total of 13,038 30-s PPG samples (5850 for SR and 7188 for AF) from this population, using CART. Among the 100 participants, 81 had persistent AF and the other 19 had long-standing persistent AF. In 15 participants, cardioversion was unsuccessful. The mean pulse rate was higher in AF samples than in SR samples (63.5, SD 9.9 vs 59.6, SD 9.9; $P<.001$). There were no adverse events or safety issues during the study.

Table 1. Baseline characteristics of the study population (N=100).

Characteristic	Value[a]
Demographics	
Age (years)	63.8 (8.5)
Male	81 (81.0)
Median body mass index (kg/m^2)	25.3 (23.5-27.1)
Median CHA_2DS_2-VASc score	2 (1-3)
Atrial fibrillation ablation history	7 (7.0)
Types of AF	
Persistent[b]	81 (81.0)
Long-standing persistent[c]	19 (19.0)
Comorbidity	
Congestive heart failure	15 (15.0)
Hypertension	57 (57.0)
Diabetes mellitus	27 (27.0)
Stroke or transient ischemic attack	4 (4.0)
Myocardial infarction or ischemic heart disease	6 (6.0)
Valvular heart disease	3 (3.0)
Dyslipidemia	35 (35.0)
Chronic renal failure	3 (3.0)
Chronic obstructive pulmonary disease	1 (1.0)
Hyperthyroidism	3 (3.0)
Antiarrhythmic agents	
Propafenone	17 (17.0)
Flecainide	10 (10.0)
Pilsicainide	3 (3.0)
Sotalol	0 (0)
Amiodarone	64 (64.0)
Beta-blocker	24 (24.0)
Nondihydropyridine calcium channel blocker	27 (27.0)
Digoxin	2 (2.0)
Anticoagulants	
Warfarin	9 (9.0)
Nonvitamin K oral anticoagulant	91 (91.0)
Other medications	
Angiotensin-converting enzyme inhibitor	100 (100)
Angiotensin II receptor blocker	29 (29.0)
Diuretics	15 (15.0)
Statin	32 (32.0)

[a]Values are mean (SD) for age, median (IQR) for body mass index and CHA_2DS_2-VASc score, or n (%) for other variables.

[b]Atrial fibrillation history for more than 1 month but less than 1 year.

[c]Atrial fibrillation history for more than 1 year.

Diagnostic Performance According to the Algorithms

The performance of CART according to the algorithms is presented in Table 2 and Figure 2. Combined with the CNN algorithm, it showed the highest performance for all the diagnostic parameters, with diagnostic accuracy of 96.89%, sensitivity of 98.96%, specificity of 94.34%, PPV of 95.55%, NPV of 98.67%, and AUC (95% CI) of 0.993 (0.992-0.993). Among the nondeep learning algorithms, SVM with the ensemble method had the highest results for all the parameters, except sensitivity and NPV, with diagnostic accuracy of 91.49%, sensitivity of 91.29%, specificity of 91.74%, PPV of 93.14%, NPV of 89.55%, and AUC (95% CI) of 0.983 (0.982-0.983). Adding RMSSD and ShE to autocorrelation as features did not significantly improve the performance of SVM (Figure 2).

Table 2. Diagnostic performance of the ring according to algorithms.

Algorithm	Accuracy, mean percentage	Sensitivity, mean percentage	Specificity, mean percentage	Positive-predictive value, mean percentage	Negative-predictive value, mean percentage	AUC[a] (95% CI)
Convolutional neural network	96.89	98.96	94.34	95.55	98.67	0.993 (0.992-0.993)
SVM[b], ensemble[c]	91.49	91.29	91.74	93.14	89.55	0.983 (0.982-0.983)
SVM, autocorrelation[d]	91.37	92.15	90.4	92.18	90.36	0.982 (0.981-0.982)
SVM, RMSSD[e]+ShE[f] [g]	84.11	90.65	76.07	82.31	86.88	0.887 (0.885-0.889)

[a]AUC: area under the receiver operating characteristic curve.

[b]SVM: support vector machine.

[c]SVM with autocorrelation, RMSSD, and ShE as features.

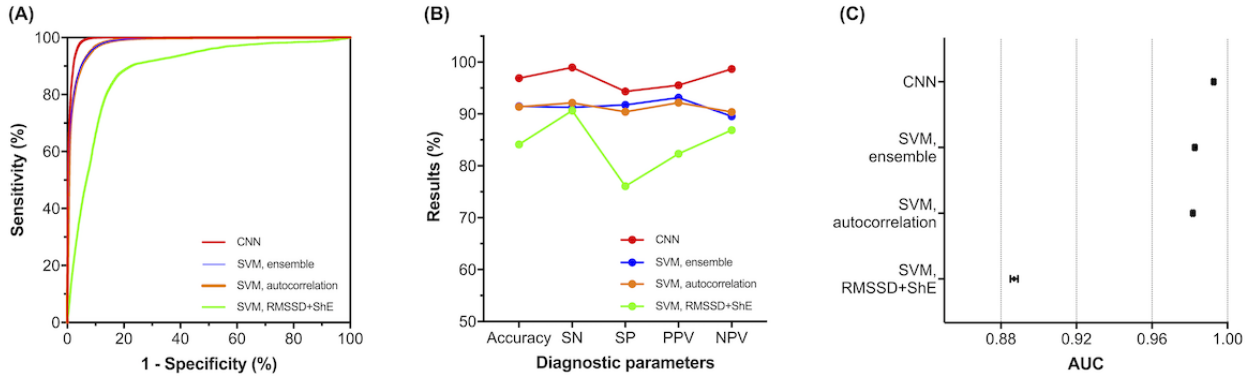
[d]SVM with autocorrelation as a feature.

[e]RMSSD: root mean square of the successive differences of RR intervals.

[f]ShE: Shannon entropy.

[g]SVM with RMSSD and ShE as features.

Figure 2. Diagnostic performance of CardioTracker (CART) according to the algorithms. CART with the deep learning algorithm achieved the highest results for all diagnostic parameters. (A) ROC curves, (B) Diagnostic parameters, and (C) AUCs according to the algorithms. AUC: area under the curve, CNN: convolutional neural network, NPV: negative-predictive value, PPV: positive-predictive value, ROC: receiver operating characteristic, SN: sensitivity, SP: specificity, SVM, autocorrelation: support vector machine with autocorrelation as a feature, SVM, RMSSD+ShE: support vector machine with root mean square of the successive differences of RR intervals and Shannon entropy as features, SVM, ensemble: support vector machine with all three features.



Impact of Sample Length

The association between sample length and the diagnostic performance of CART is presented in Figure 3 and Table 3. The figure only presents the results of CART with CNN. As expected, all diagnostic parameters decreased as sample length shortened. Using 10-s PPG segments, CART achieved diagnostic accuracy of 94.72%, sensitivity of 97.46%, specificity of 91.35%, PPV of 93.26%, NPV of 96.69%, and AUC (95% CI) of 0.985 (0.985-0.986).

Figure 3. The diagnostic performance of CardioTracker according to sample length. In general, longer lengths of photoplethysmography samples had higher diagnostic performances. AUC: area under the curve, NPV: negative-predictive value, PPV: positive-predictive value, SN: sensitivity, SP: specificity.

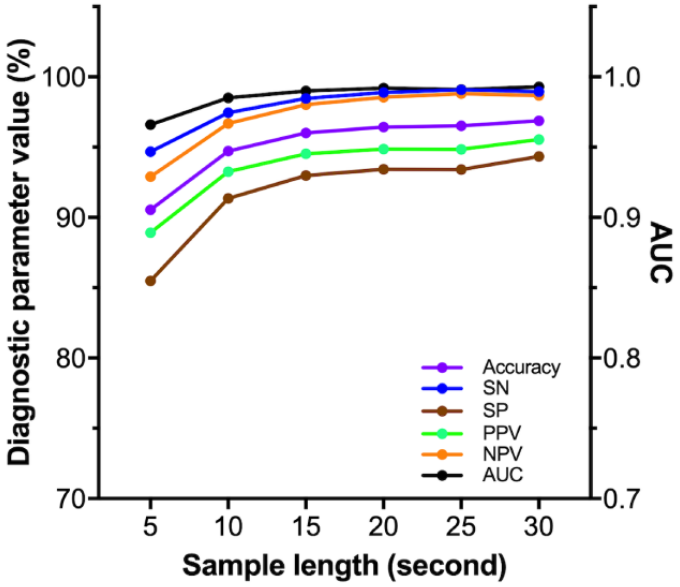


Table 3. Diagnostic performance according to sample length.

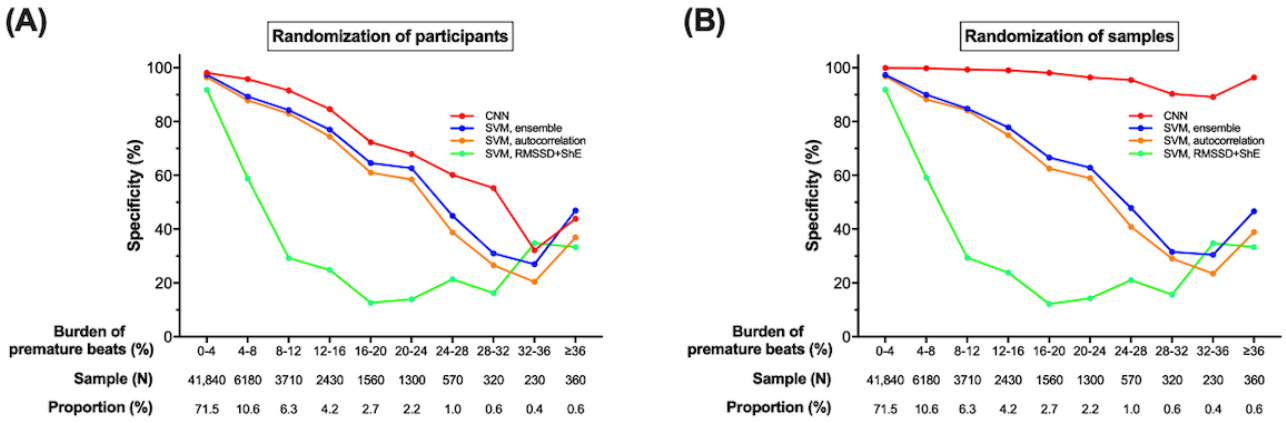
Duration (s)	Accuracy, mean percentage	Sensitivity, mean percentage	Specificity, mean percentage	Positive-predictive value, mean percentage	Negative-predictive value, mean percentage	AUC[a] (95% CI)
30	96.89	98.96	94.34	95.55	98.67	0.993 (0.992-0.993)
25	96.53	99.08	93.40	94.85	98.80	0.991 (0.990-0.991)
20	96.44	98.89	93.43	94.87	98.56	0.992 (0.992-0.992)
15	96.01	98.47	92.99	94.52	98.02	0.990 (0.990-0.990)
10	94.72	97.46	91.35	93.26	96.69	0.985 (0.985-0.986)
5	90.55	94.68	85.49	88.91	92.90	0.966 (0.965-0.966)

[a]AUC: area under the receiver operating characteristic curve.

Impact of Premature Beats

Each SR-labelled PPG sample may have a record of atrial or ventricular premature beats, which can be confirmed by evaluating synchronized ECG. We evaluated the specificity of our device by assessing the burden of premature beats (Figure 4). The total number of samples was 10 times the number of SR samples (58,500), as the validation processes were repeated 10 times. When participants were randomized for the five-fold cross-validation processes such that the algorithm always encountered new participants in the testing, higher burdens of premature beats deteriorated the specificity of CART, regardless of the algorithm. Among the algorithms, CNN maintained the highest results for most cases of premature beat burdens. When samples were randomized such that the algorithm might encounter the same participants in testing, there was an improvement in CNN performance, especially for higher burdens of premature beats, and CNN maintained overall consistent results, regardless of the burden. As this validation process simulates a situation with a sufficiently large number of participants in the training dataset, this finding implies that the performance would improve with an increasing population, regardless of the burden of premature beats.

Figure 4. The specificity of CardioTracker according to the burden of premature beats. (A) The five-fold cross-validation process with randomization of participants. There was a decreasing trend of specificity according to increasing burden of premature beats. However, the convolutional neural network (CNN) maintained the highest results for most cases. (B) The five-fold cross-validation process with randomization of samples. The CNN improved specificity in especially high burden of premature beats. SVM, autocorrelation: support vector machine with autocorrelation as a feature, SVM, RMSSD+ShE: support vector machine with root mean square of the successive differences of RR intervals and Shannon entropy as features, SVM, ensemble: support vector machine with all three features.

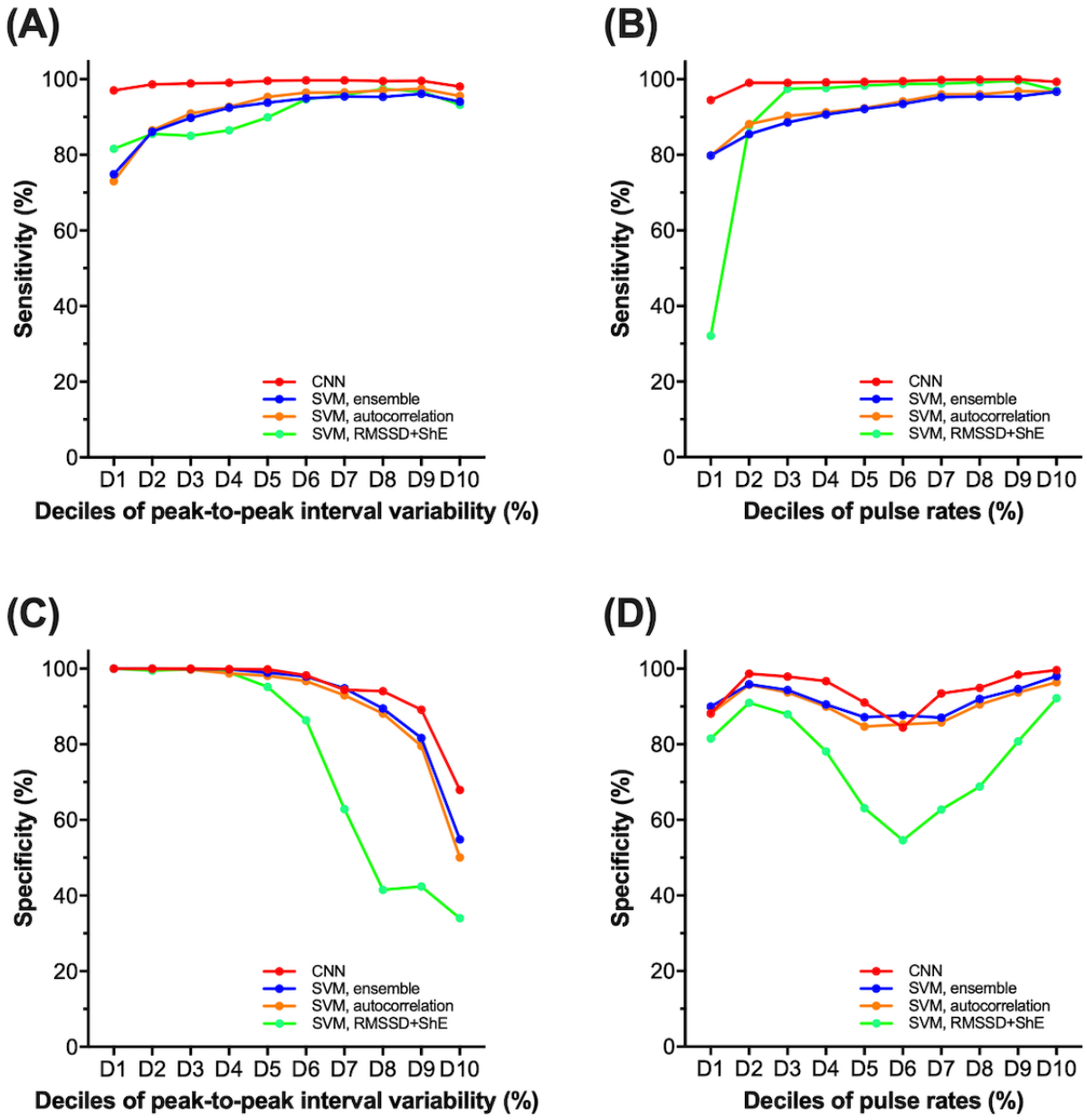


Impact of the Variability of Peak-to-Peak Intervals and the Pulse Rate

The performance of CART can be affected by the characteristics of the PPG samples. We evaluated whether the variability of peak-to-peak intervals and the pulse rate affected the performance (Figure 5). For sensitivity, higher peak-to-peak interval variability and faster pulse rates were associated with higher sensitivity for SVM. However, for CNN, neither peak-to-peak interval variability nor pulse rate had such a relevant association with sensitivity. This finding suggests that CART with a deep learning algorithm is less affected by peak-to-peak interval variability or the pulse rate of AF.

For specificity, the performance of CART decreased with higher peak-to-peak interval variability regardless of the algorithm. However, for CNN, only the extremes of the variability (the ninth and the tenth deciles) had decreased specificity less than 90%. This finding was expected, as SR with higher peak-to-peak interval variability mimics AF to a great extent. There was a nonlinear association between specificity and pulse rate, and in general, the results were the highest with CNN. The complicated association between specificity and pulse rate can be mostly explained by evaluating the association between the burden of premature beats and the pulse rate (Multimedia Appendix 4). For example, lower specificity for the sixth decile of the pulse rate can be due to the higher burden of premature beats.

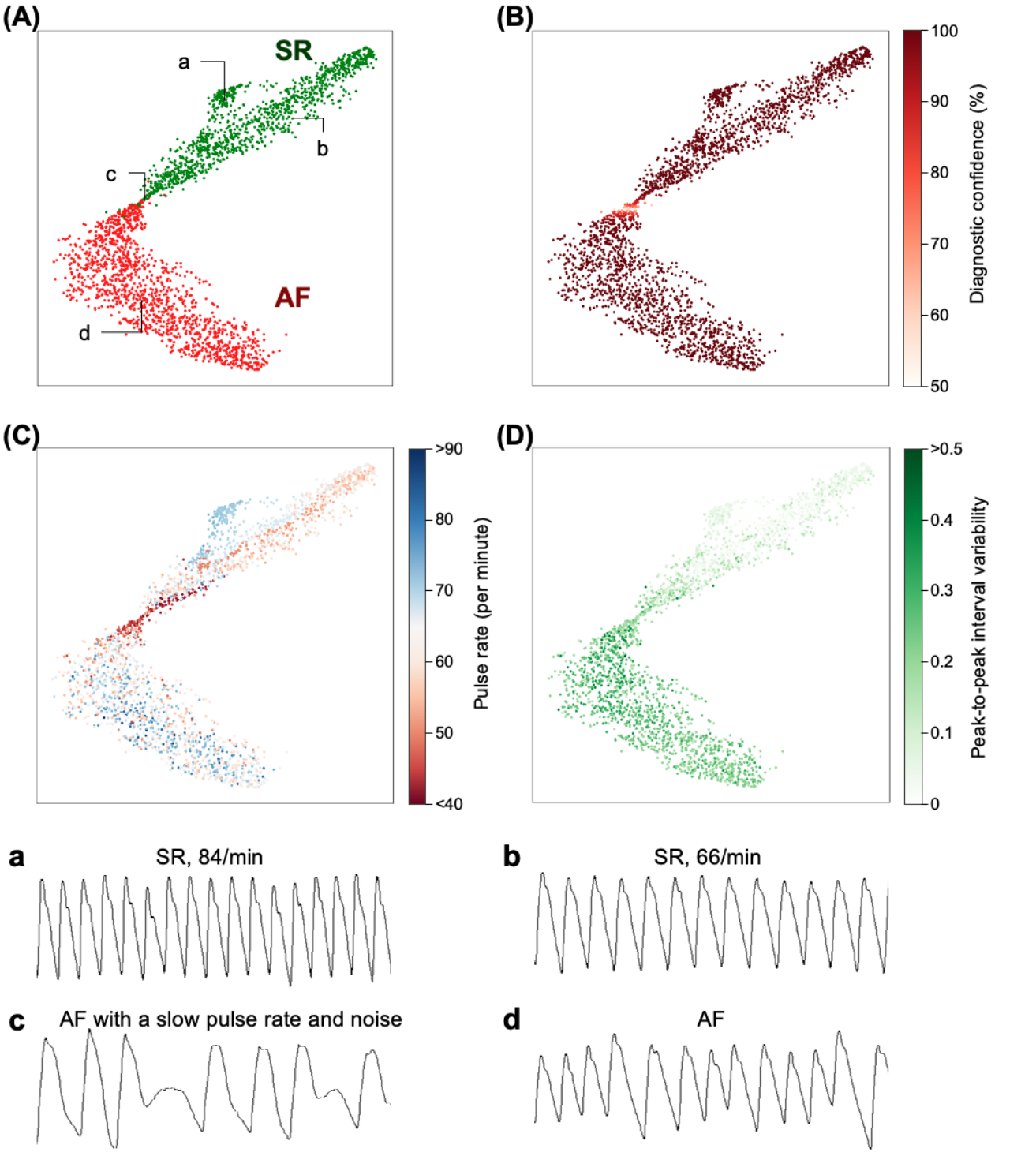
Figure 5. The sensitivity and specificity of CardioTracker according to the characteristics of samples. (A) and (B) With the deep learning algorithm, there were no definite associations between the sensitivity and peak-to-peak interval variability or the pulse rate. (C) The specificity generally decreased with higher peak-to-peak interval variability. (D) There was generally a U-shape association between specificity and the pulse rate. CNN: convolutional neural network, SVM, autocorrelation: support vector machine with autocorrelation as a feature, SVM, RMSSD+ShE: support vector machine with root mean square of the successive differences of RR intervals and Shannon entropy as features, SVM, ensemble: support vector machine with all three features.



Visualization of Deep Learning Analyses

The deep learning analyses for CART are illustrated in Figure 6 by mapping extracted features from the deep learning algorithm into two-dimensional space. According to the t-distributed stochastic neighbor embedding plot, the cluster of AF was well differentiated from the counterpart of SR. In the region where the two clusters overlapped, lower diagnostic confidences were observed, which suggests that the deep learning algorithm mostly failed when the PPG samples belonged to this region. When we applied heatmaps with the pulse rate and peak-to-peak interval variability, this region had characteristics with lower pulse rates and modest variabilities. When we inspected the actual PPG data, this region also exhibited noisy signals. Therefore, PPG samples with lower pulse rates, modest variabilities, and noise would likely have low diagnostic performance. The cluster of AF was homogeneous in terms of the pulse rate, whereas the cluster of SR had distinctive subportions according to the actual pulse rate.

Figure 6. Visualization of deep learning analyses. The deep learning analyses of CardioTracker are plotted with the t-SNE method. The upper panel: (A) The two clusters of AF and SR were well differentiated from each other, leaving a small overlapped potion. (B), (C), and (D) The overlapped region showed low diagnostic confidence, low pulse rates, and modest peak-to-peak interval variability. The lower panel: typical examples of photoplethysmography samples. AF: atrial fibrillation, SR: sinus rhythm, t-SNE: t-distributed stochastic neighbor embedding.



CardioTracker and the Conventional Pulse Oximeter

We evaluated the two index tests in parallel and observed the changes in diagnostic performances according to the threshold level of diagnostic confidence. Compared with the conventional pulse oximeter, CART showed comparable diagnostic performance (Table 4). In both devices, all the diagnostic parameters improved with increasing threshold levels (Multimedia Appendix 5). This finding is expected, as the diagnosis would become more accurate for samples with higher diagnostic confidence. However, this improvement was counter-balanced by increasing the proportion of filtered samples (not tested by the deep learning algorithm). From these findings, CART appears to be comparable to the conventional pulse oximeter when used as a PPG measurement device. Moreover, the performance of CART can be tuned by applying different threshold levels.

Table 4. Comparison of diagnostic performance between CardioTracker and the conventional pulse oximeter at the fingertip (control).

Device	Accuracy, mean percentage	Sensitivity, mean percentage	Specificity, mean percentage	Positive-predictive value, mean percentage	Negative-predictive value, mean percentage	AUC[a] (95% CI)[b]
CART[c]	96.89	98.96	94.34	95.55	98.67	0.993 (0.992-0.993)
Control	97.50	99.66	94.89	95.95	99.56	0.995 (0.995-0.995)

[a]AUC: area under the receiver operating characteristic curve.

[b]The standard error by the binomial exact test was less than 0.01.

[c]CART: CardioTracker.

External Validation of CardioTracker in Healthy Subjects

A total of 1740 PPG samples were obtained from 20 healthy subjects with SR. Among these samples with repeated testing 10 times, 94.99% (16,529/17,400) of the testing cases were correctly identified with SR.

Discussion

Principal Findings

This prospective observational cohort study evaluated the diagnostic performance of a ring-type wearable device (CART) to detect AF. To the best of our knowledge, this is the first clinical study to analyze the performance of a ring-type wearable device designed for detecting AF with PPG. The study had several findings. First, we found that the deep learning algorithm can maximize the performance of CART solely based on PPG. Second, a PPG measurement period of about 10 s may be sufficient to detect AF. Third, the data from a sufficiently large number of participants may further improve the performance of CART by enhancing the deep learning process, especially for difficult cases in which the high burden of premature beats mimics AF. Fourth, among the diagnostic parameters, sensitivity may be maintained at a consistently high level regardless of the variability of peak-to-peak intervals or the pulse rate of PPG signals. Fifth, although CART measures PPG at the middle of the finger, which is not the location where a conventional pulse oximeter measures the impulse, its performance is comparable to that of a conventional device. Lastly, we performed external validation of CART with the deep learning algorithm in healthy subjects and observed that the CNN algorithm can diagnose SR accurately.

Screening for Atrial Fibrillation in a High-Risk Population

AF is known to cause about 10% of the total cases of stroke, and it has been shown to increase the risk of stroke even when discovered incidentally through screening [26]. AF screening is beneficial in an appropriate setting for patients in all localities. Therefore, early diagnosis of AF with appropriate anticoagulant therapy is expected to reduce the risk of ischemic stroke. However, early diagnosis is challenging because paroxysmal or asymptomatic episodes are common. Therefore, further research is needed to find more convenient and effective screening methods. Based on this aspect, PPG has recently attracted attention as a method of AF screening because it can be continuously monitored with appropriate equipment and its measurement is convenient.

Utility of Photoplethysmography to Detect Atrial Fibrillation

This study used PPG signals measured from CART to detect AF. As there is a good correlation between each pulse of PPG and the corresponding QRS complex on ECG, it is feasible to diagnose AF with PPG. Considering that PPG has limited capability to detect atrial electrical activity, many PPG algorithms have been studied to detect AF using the randomness of peak-to-peak intervals [27,28]. Two issues should be addressed. First, there is uncertainty as to the choice of algorithm used to detect AF. For detecting AF with PPG, deep learning algorithms have been known to achieve the highest diagnostic results so far [14]. The possible reason is that nondeep learning algorithms use only specific data features, which are invented by humans, whereas deep learning algorithms analyze the entire dataset without human guidance. Second, the optimal anatomical location for PPG measurement is debatable. Multiple studies have evaluated the diagnostic value of PPG measured at various sites, including the wrist and face [8,10,12]. However, a recent study showed that the finger has the highest quality of PPG signals [16]. Therefore, to diagnose AF more effectively, it is probably best to analyze PPG signals from the finger. In summary, the most effective AF diagnosis is possible when analyzing PPG signals from the finger and using deep learning algorithms.

Wearable Devices to Detect Atrial Fibrillation With Photoplethysmography

The method for screening AF should be not only supported by sufficient diagnostic precision but also convenient for patients. The strategy of carrying a portable device involving point-of-care testing is not only inconvenient but also ineffective in that it can miss the diagnosis when AF is asymptomatic. In this context, a wearable device that continuously monitors PPG signals in the background without user intervention would be easy to use and efficient for diagnosis. If CART is worn all day, PPG signals can be continuously monitored; therefore, more AF episodes could be detected. However, continuous monitoring might increase the chance of collecting other signals, such as noise caused by movements in daily life, leading to a decrease in the accuracy of diagnosis. To resolve this problem, we need engineering technology for estimating and correcting motion artifacts through the use of accelerometer sensors in addition to PPG signal analysis. Second, various atrial tachyarrhythmia episodes other than AF could be detected more frequently, which might lower the diagnostic performance for AF detection by

CART. Sufficient data should be collected for various atrial tachyarrhythmias in addition to AF to improve machine learning analyses.

Besides, when PPG signals are continuously monitored, various atrial tachyarrhythmia episodes can occur in addition to AF, which may lower the diagnostic performance of CART. If atrial tachyarrhythmia episodes occur frequently, the irregularity of the PPG signal is expected to be similar to that for AF, and in this case, the diagnostic performance of CART may deteriorate. To solve this issue, sufficient data should be collected for various arrhythmias in addition to AF to improve machine learning analyses.

Although wearable devices may not have become popular due to their availability and cost-effectiveness, the usefulness of such devices in the detection of AF has been studied [29]. The wrist-type device is one of the most widely studied wearable devices [13,30,31]. However, the WATCH AF Trial also reported that a high proportion (22%) of PPG signals from smartwatches had an insufficient signal quality for evaluation [31]. Therefore, measuring PPG signals on the wrist may lead to poor signal quality. In order to resolve this issue, other types of wearable devices are necessary, and the ring-type device, which measures PPG signals from the finger, might be ideal. Therefore, a ring-type wearable device could be useful as a new diagnostic tool for high-risk populations in the future.

Limitations

There are some limitations in this study. First, noise in PPG signals, such as motion artifacts, might affect the analysis. However, motion artifacts were minimized as every participant was required to lie still on the bed during the PPG measurement. The diagnostic performance of CART in an ambulatory setting will be tested, but a sophisticated deep learning algorithm should be developed beforehand to deal with motion artifacts. Second, the performance of CART was not assessed for other arrhythmias. Future studies should analyze the diagnosis of other arrhythmias using PPG signals. Third, the duration of monitoring was relatively short. Longer monitoring times would allow further deep learning training and subsequently yield better results than our results. Fourth, we did not compare performance between CART and other commercially available wrist-type wearable devices. Further studies will provide insights into this issue. Fifth, economic assessment of CART cannot be performed yet. However, in the case of AliveCor, it was shown that a wearable device could be cost-effective for AF screening [29]. Likewise, CART is also expected to reduce the economic burden of diagnosing AF if its market price is reasonable. This economic evaluation requires further research. Sixth, since a 20-s overlap existed between consecutive samples obtained from a subject during data augmentation, it is possible that even if different samples existed in the training and testing datasets, some sections were the same, and thus, the diagnostic performance was improved. Seventh, AF diagnosis by PPG only has limitations. Adding the on-demand recording function of single-lead ECG, similar to an Apple Watch, to CART may compensate for the limitations that arise in diagnoses based on PPG signals. In this case, if AF is suspected during PPG monitoring, a notification can be sent to the user to check the electrocardiogram, so that the user can more clearly check for AF. The validation of such a function would be performed in future research. Lastly, even though there were no adverse events of CART, potential safety issues in long-term use should be addressed in a subsequent study.

Conclusions

In this study, we validated the performance of a ring-type wearable device (CART) to diagnose AF using PPG signals. The deep learning algorithm aimed to analyze PPG rhythms and suggested a dichotomous diagnosis of either AF or SR. CART with deep learning analysis of PPG signals had good diagnostic performance without relying on ECG. Moreover, as a PPG measurement device, CART generated results comparable to those of a conventional medical-grade pulse oximeter. This new device may be promising for the detection of AF in high-risk or asymptomatic populations.

Acknowledgments

We would like to thank Dajeong Heo, who helped in the photoplethysmography measurement processes during the study. This work was supported by Sky Labs Inc, Seongnam, Republic of Korea, and by the grant no. 0320202040 from the Seoul National University Hospital Research Fund.

Authors' Contributions

SK and JH have made equal contributions to the design of the study, acquisition, analysis, and interpretation of data, creation of figures and tables, and draft of the manuscript. EKC and YY contributed equally in correspondence to the study and revision of the manuscript.

Conflicts of Interest

SK, JH, EL, and SO: None declared. EKC, ERJ, BKK, and YY: Stockholders of Sky Labs Inc, Seongnam, Republic of Korea. BL and CB: Employees of Sky Labs Inc.

Multimedia Appendix 1
Study flowchart. AF: atrial fibrillation, CNN: convolutional neural network, CV: cardioversion, ECG: electrocardiography, PPG: photoplethysmography, SVM: support vector machine.

[PNG File , 233 KB - jmir_v22i5e16443_app1.png]

Multimedia Appendix 2
The first row of the figure shows the 15-min measurements from a subject with the single-lead ECG, the PPG from the conventional pulse oximeter, and the PPG from CardioTracker. The second row of the figure shows one of the 30-s fractions of the raw data. The third row of the figure shows the preprocessed 30-s sample of the PPG signals obtained from the two devices. ECG: electrocardiography, PPG: photoplethysmography.
[PNG File , 4835 KB - jmir_v22i5e16443_app2.png]

Multimedia Appendix 3
The left figure illustrates CardioTracker (CART) and its wireless charging station. The right figure shows the position of the light-emitting diode (LED) and photodiode (PD) of CART.
[PNG File , 824 KB - jmir_v22i5e16443_app3.png]

Multimedia Appendix 4
The burden of premature beats according to the deciles of the pulse rate. Each error bar represents the 95% CI of the corresponding burden of premature beats.
[PNG File , 179 KB - jmir_v22i5e16443_app4.png]

Multimedia Appendix 5
The performances of CardioTracker and a conventional pulse oximeter according to threshold levels of diagnostic probability. Both devices showed improved performance with increasing threshold levels. NPV: negative-predictive value, PPV: positive-predictive value.
[PNG File , 231 KB - jmir_v22i5e16443_app5.png]

References

1. Lee S, Choi E, Han K, Cha M, Oh S. Trends in the incidence and prevalence of atrial fibrillation and estimated thromboembolic risk using the CHADS-VASc score in the entire Korean population. Int J Cardiol 2017 Jun 01;236:226-231. [doi: 10.1016/j.ijcard.2017.02.039] [Medline: 28233629]
2. Krijthe BP, Kunst A, Benjamin EJ, Lip GY, Franco OH, Hofman A, et al. Projections on the number of individuals with atrial fibrillation in the European Union, from 2000 to 2060. Eur Heart J 2013 Sep;34(35):2746-2751 [FREE Full text] [doi: 10.1093/eurheartj/eht280] [Medline: 23900699]
3. Lee H, Kim T, Baek Y, Uhm J, Pak H, Lee M, et al. The Trends of Atrial Fibrillation-Related Hospital Visit and Cost, Treatment Pattern and Mortality in Korea: 10-Year Nationwide Sample Cohort Data. Korean Circ J 2017 Jan;47(1):56-64 [FREE Full text] [doi: 10.4070/kcj.2016.0045] [Medline: 28154592]
4. Wyse DG, Van Gelder IC, Ellinor PT, Go AS, Kalman JM, Narayan SM, et al. Lone atrial fibrillation: does it exist? J Am Coll Cardiol 2014 May 06;63(17):1715-1723 [FREE Full text] [doi: 10.1016/j.jacc.2014.01.023] [Medline: 24530673]
5. Heidt ST, Kratz A, Najarian K, Hassett AL, Oral H, Gonzalez R, et al. Symptoms In Atrial Fibrillation: A Contemporary Review And Future Directions. J Atr Fibrillation 2016;9(1):1422 [FREE Full text] [doi: 10.4022/jafib.1422] [Medline: 27909518]
6. Jonas DE, Kahwati LC, Yun JD, Middleton JC, Coker-Schwimmer M, Asher GN. Screening for Atrial Fibrillation With Electrocardiography: Evidence Report and Systematic Review for the US Preventive Services Task Force. JAMA 2018 Aug 07;320(5):485-498. [doi: 10.1001/jama.2018.4190] [Medline: 30088015]
7. Camm AJ. The Role of Continuous Monitoring in Atrial Fibrillation Management. Arrhythm Electrophysiol Rev 2014 May;3(1):48-50 [FREE Full text] [doi: 10.15420/aer.2011.3.1.48] [Medline: 26835066]
8. Yan BP, Lai WH, Chan CK, Chan SC, Chan L, Lam K, et al. Contact-Free Screening of Atrial Fibrillation by a Smartphone Using Facial Pulsatile Photoplethysmographic Signals. J Am Heart Assoc 2018 Apr 05;7(8) [FREE Full text] [doi: 10.1161/JAHA.118.008585] [Medline: 29622592]
9. Chan P, Wong C, Poh YC, Pun L, Leung WW, Wong Y, et al. Diagnostic Performance of a Smartphone-Based Photoplethysmographic Application for Atrial Fibrillation Screening in a Primary Care Setting. J Am Heart Assoc 2016 Jul 21;5(7) [FREE Full text] [doi: 10.1161/JAHA.116.003428] [Medline: 27444506]
10. Conroy T, Guzman JH, Hall B, Tsouri G, Couderc J. Detection of atrial fibrillation using an earlobe photoplethysmographic sensor. Physiol Meas 2017 Sep 26;38(10):1906-1918. [doi: 10.1088/1361-6579/aa8830] [Medline: 28836507]
11. Mc MD, Chong JW, Soni A, Saczynski JS, Esa N, Napolitano C, et al. PULSE-SMART: Pulse-Based Arrhythmia Discrimination Using a Novel Smartphone Application. J Cardiovasc Electrophysiol 2016 Jan;27(1):51-57 [FREE Full text] [doi: 10.1111/jce.12842] [Medline: 26391728]

12. Bonomi AG, Schipper F, Eerikäinen LM, Margarito J, van Dinther R, Muesch G, et al. Atrial Fibrillation Detection Using a Novel Cardiac Ambulatory Monitor Based on Photo-Plethysmography at the Wrist. J Am Heart Assoc 2018 Aug 07;7(15):e009351 [FREE Full text] [doi: 10.1161/JAHA.118.009351] [Medline: 30371247]
13. Perez MV, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T, Apple Heart Study Investigators. Large-Scale Assessment of a Smartwatch to Identify Atrial Fibrillation. N Engl J Med 2019 Nov 14;381(20):1909-1917. [doi: 10.1056/NEJMoa1901183] [Medline: 31722151]
14. Kwon S, Hong J, Choi E, Lee E, Hostallero DE, Kang WJ, et al. Deep Learning Approaches to Detect Atrial Fibrillation Using Photoplethysmographic Signals: Algorithms Development Study. JMIR Mhealth Uhealth 2019 Jun 06;7(6):e12770 [FREE Full text] [doi: 10.2196/12770] [Medline: 31199302]
15. Nilsson L, Goscinski T, Kalman S, Lindberg L, Johansson A. Combined photoplethysmographic monitoring of respiration rate and pulse: a comparison between different measurement sites in spontaneously breathing subjects. Acta Anaesthesiol Scand 2007 Oct;51(9):1250-1257. [doi: 10.1111/j.1399-6576.2007.01375.x] [Medline: 17711563]
16. Hartmann V, Liu H, Chen F, Qiu Q, Hughes S, Zheng D. Quantitative Comparison of Photoplethysmographic Waveform Characteristics: Effect of Measurement Site. Front Physiol 2019;10:198 [FREE Full text] [doi: 10.3389/fphys.2019.00198] [Medline: 30890959]
17. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. : Curran Associates Inc; 2012 Presented at: Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1; December 3-8, 2012; Lake Tahoe, Nevada p. 1097-1106. [doi: 10.1145/3065386]
18. Tamura T, Maeda Y, Sekine M, Yoshida M. Wearable Photoplethysmographic Sensors—Past and Present. Electronics 2014 Apr 23;3(2):282-302. [doi: 10.3390/electronics3020282]
19. Zhang X, Zhou X, Lin M, Sun J. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2018:6848-6856. [doi: 10.1109/CVPR.2018.00716]
20. Kingma DP, Ba JL. Adam: A method for stochastic optimization. arXiv:14126980 2014.
21. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research 2014;15(56):1929-1958.
22. Guo C, Pleiss G, Sun Y, Weinberger KQ. On Calibration of Modern Neural Networks. arXiv:170604599 2017. [doi: 10.1090/mbk/121/79]
23. Suykens J, Vandewalle J. Least Squares Support Vector Machine Classifiers. Neural Processing Letters 1999;9:293-300. [doi: 10.1023/A:1018628609742]
24. Poh M, Poh YC, Chan P, Wong C, Pun L, Leung WW, et al. Diagnostic assessment of a deep learning system for detecting atrial fibrillation in pulse waveforms. Heart 2018 Dec;104(23):1921-1928. [doi: 10.1136/heartjnl-2018-313147] [Medline: 29853485]
25. Lee K, Choi HO, Min SD, Lee J, Gupta BB, Nam Y. A Comparative Evaluation of Atrial Fibrillation Detection Methods in Koreans Based on Optical Recordings Using a Smartphone. IEEE Access 2017;5:11437-11443. [doi: 10.1109/ACCESS.2017.2700488]
26. Freedman B, Camm J, Calkins H, Healey JS, Rosenqvist M, Wang J, AF-Screen Collaborators. Screening for Atrial Fibrillation: A Report of the AF-SCREEN International Collaboration. Circulation 2017 May 09;135(19):1851-1867. [doi: 10.1161/CIRCULATIONAHA.116.026693] [Medline: 28483832]
27. Shan SM, Tang SC, Huang PW, Lin YM, Huang WH, Lai DM, et al. Reliable PPG-based algorithm in atrial fibrillation detection. : IEEE; 2016 Presented at: IEEE Biomedical Circuits and Systems Conference (BioCAS); October 17-19, 2016; Shanghai p. 340-343. [doi: 10.1109/BioCAS.2016.7833801]
28. Chong JW, Esa N, McManus DD, Chon KH. Arrhythmia discrimination using a smart phone. IEEE J Biomed Health Inform 2015 May;19(3):815-824 [FREE Full text] [doi: 10.1109/JBHI.2015.2418195] [Medline: 25838530]
29. Giebel GD, Gissel C. Accuracy of mHealth Devices for Atrial Fibrillation Screening: Systematic Review. JMIR Mhealth Uhealth 2019 Jun 16;7(6):e13641 [FREE Full text] [doi: 10.2196/13641] [Medline: 31199337]
30. Tison GH, Sanchez JM, Ballinger B, Singh A, Olgin JE, Pletcher MJ, et al. Passive Detection of Atrial Fibrillation Using a Commercially Available Smartwatch. JAMA Cardiol 2018 May 01;3(5):409-416 [FREE Full text] [doi: 10.1001/jamacardio.2018.0136] [Medline: 29562087]
31. Dörr M, Nohturfft V, Brasier N, Bosshard E, Djurdjevic A, Gross S, et al. The WATCH AF Trial: SmartWATCHes for Detection of Atrial Fibrillation. JACC Clin Electrophysiol 2019 Feb;5(2):199-208 [FREE Full text] [doi: 10.1016/j.jacep.2018.10.006] [Medline: 30784691]

Abbreviations

AF: atrial fibrillation
AUC: area under the curve
CART: CardioTracker
CNN: convolutional neural network

ECG: electrocardiography
NPV: negative-predictive value
PPG: photoplethysmography
PPV: positive-predictive value
RMSSD: root mean square of the successive differences of RR intervals
ShE: Shannon entropy
SR: sinus rhythm
SVM: support vector machine

Edited by G Eysenbach; submitted 30.09.19; peer-reviewed by C Gissel, J Parak, A Vehkaoja; comments to author 26.11.19; revised version received 24.02.20; accepted 21.03.20; published 21.05.20.

Please cite as:
Kwon S, Hong J, Choi EK, Lee B, Baik C, Lee E, Jeong ER, Koo BK, Oh S, Yi Y
Detection of Atrial Fibrillation Using a Ring-Type Wearable Device (CardioTracker) and Deep Learning Analysis of Photoplethysmography Signals: Prospective Observational Proof-of-Concept Study
J Med Internet Res 2020;22(5):e16443
URL: http://www.jmir.org/2020/5/e16443/
doi:10.2196/16443
PMID:32348254

Original Paper

What Medical Crowdfunding Campaigns Can Tell Us About Local Health System Gaps and Deficiencies: Exploratory Analysis of British Columbia, Canada

Jeremy Snyder[1], PhD; Marco Zenone[1], BA; Valorie Crooks[2], PhD; Nadine Schuurman[2], PhD

[1]Faculty of Health Sciences, Simon Fraser University, Burnaby, BC, Canada

[2]Department of Geography, Simon Fraser University, Burnaby, BC, Canada

Corresponding Author:
Jeremy Snyder, PhD
Faculty of Health Sciences
Simon Fraser University
Blusson Hall, Room 11300
8888 University Drive
Burnaby, BC, V5A 1S6
Canada
Phone: 1 7787823258
Email: jcs12@sfu.ca

Abstract

Background: There are a range of perceived gaps and shortcomings in the publicly funded Canadian health system. These include wait times for care, lack of public insurance coverage for dental care and pharmaceuticals, and difficulties accessing specialist care. Medical crowdfunding is a response to these gaps where individuals raise funds from their social networks to address health-related needs.

Objective: This study aimed to investigate the potential of crowdfunding data to better understand what health-related needs individuals are using crowdfunding for, how these needs compare with the existing commentary on health system deficiencies, and the advantages and limitations of using crowdfunding campaigns to enhance or augment our understanding of perceived health system deficiencies.

Methods: Crowdfunding campaigns were scraped from the GoFundMe website. These campaigns were then limited to those originating in the metropolitan Vancouver region of two health authorities during 2018. These campaigns were then further limited to those raising funds to allow the treatment of a medical problem or related to needs arising from ill health. These campaigns were then reviewed to identify the underlying health issue and motivation for pursuing crowdfunding.

Results: We identified 423 campaigns for health-related needs. These campaigns requested CAD $8,715,806 (US $6,088,078) in funding and were pledged CAD $3,477,384 (US $2,428,987) from 27,773 donors. The most common underlying medical condition for campaign recipients was cancer, followed by traumatic injuries from collisions and brain injury and stroke. By far, the most common factor of motivation for crowdfunding was seeking financial support for wages lost because of illness (232/684, 33.9%). Some campaigns (65/684, 9.5%) sought help with purchasing medical equipment and supplies; 8.2% (56/684) sought to fund complementary, alternative, or unproven treatments including experimental interventions; 7.2% (49/684) sought financial support to cover travel-related costs, including in-province and out-of-province (49/684, 7.2%) travel; and 6.3% (43/684) campaigns sought help to pay for medication.

Conclusions: This analysis demonstrates the potential of crowdfunding data to present timely and context-specific user-created insights into the perceived health-related financial needs of some Canadians. Although the literature on perceived limitations of the Canadian health system focuses on wait times for care and limited access to specialist services, among other issues, these campaigners were much more motivated by gaps in the wider social system such as costs related to unpaid time off work and travel to access care. Our findings demonstrate spatial differences in the underlying medical problems, motivations for crowdfunding, and success using crowdfunding that warrants additional attention. These differences may support established concerns that medical crowdfunding is most commonly used by individuals from relatively privileged socioeconomic backgrounds. We encourage the development of new resources to harness the power of crowdfunding data as a supplementary source of information for Canadian health system stakeholders.

(J Med Internet Res 2020;22(5):e16982) doi:10.2196/16982

KEYWORDS

crowdfunding; exploratory analysis; Canada; health system

Introduction

Background

A core aim of the Canadian health care system—and many other publicly funded health systems—is universal coverage for medically necessary services [1]. Nonetheless, Canadian health care users and policy makers report a range of perceived or experienced deficiencies in this system that impose financial costs on its users. These deficiencies include barriers to accessing care that motivate the use of private medical treatment and out-of-pocket costs when accessing publicly funded treatment [2,3].

Public reporting on systemic barriers to accessing care in Canada often focuses on wait times and health human resource shortages in key areas. A 2013 survey of Canadians found that waiting for a medical appointment and difficulty getting an appointment were the most common problems cited by those reporting access barriers. These barriers were greatest in relation to accessing specialist care, nonemergency surgery, and diagnostic testing. Canadians who were aged younger than 65 years, females, immigrants, people with some postsecondary education, and those with specific health needs were more likely to report barriers to accessing care than other groups. Geographically, Canadians in Quebec and the Western provinces, including British Columbia, were more likely to report access barriers than those living elsewhere in the country [4].

The Canadian Institute for Health Information has found that access barriers because of wait times vary by location and treatment sought. They report that 30% of Canadians perceive facing wait times for access to surgery for hip replacement or cataracts that are longer than recommended. At the same time, hip fracture repair was provided within the recommended wait times for 88% of Canadians, and 97% received timely radiation therapy [5]. British Columbia performs below the Canadian average in all wait time categories, with 67% of residents accessing hip replacement treatment within the recommended wait time, 59% for knee replacement, 85% for hip fracture repair, 64% for cataract surgery, and 93% for radiation therapy [6]. These perceived barriers to accessing care have been cited as factors motivating the expansion of private insurance in Canada and private payment for treatment domestically and abroad via medical tourism [7-9].

In addition to these perceived barriers to accessing care, several key gaps in coverage for services that are not covered under the Canada Health Act and thus are not part of the publicly funded health system have been identified [1]. These gaps include lack of or inadequate payment for prescription drugs, eye care, and dental care, which may be met by other social programs or private health insurance, depending on one's income, place of residence, and employment [10,11]. Other nonreimbursed out-of-pocket expenses encountered by Canadians include parking payments at hospitals, medical devices used at home, accommodations while accessing nonlocal care, complementary and alternative treatments, home accessibility modifications, physical rehabilitation, in-home care, and travel costs [12-15]. Limited public coverage of prescription drugs, in particular, has received significant academic, policy, and political attention [16].

Some Canadians who are facing these perceived deficiencies in the Canadian health system are turning to medical crowdfunding to assist them with raising funds to support options such as seeking treatment abroad or paying for care and services that are not covered by the health and social care systems. Medical crowdfunding is a practice whereby individuals seek funds for health-related needs from their social networks via Web-based fundraising platforms. By far, the largest medical crowdfunding platform, GoFundMe, has been growing by 300% per year in 2016 and has raised funding from 25 million donors through 2 million crowdfunding campaigns [17]. In 2017, it was reported that it had raised US $3 billion since 2010 and has been raising US $140 million per month in donations. [18] More recently, the total raised by GoFundMe was reported as US $5 billion and growing [19].

Objectives

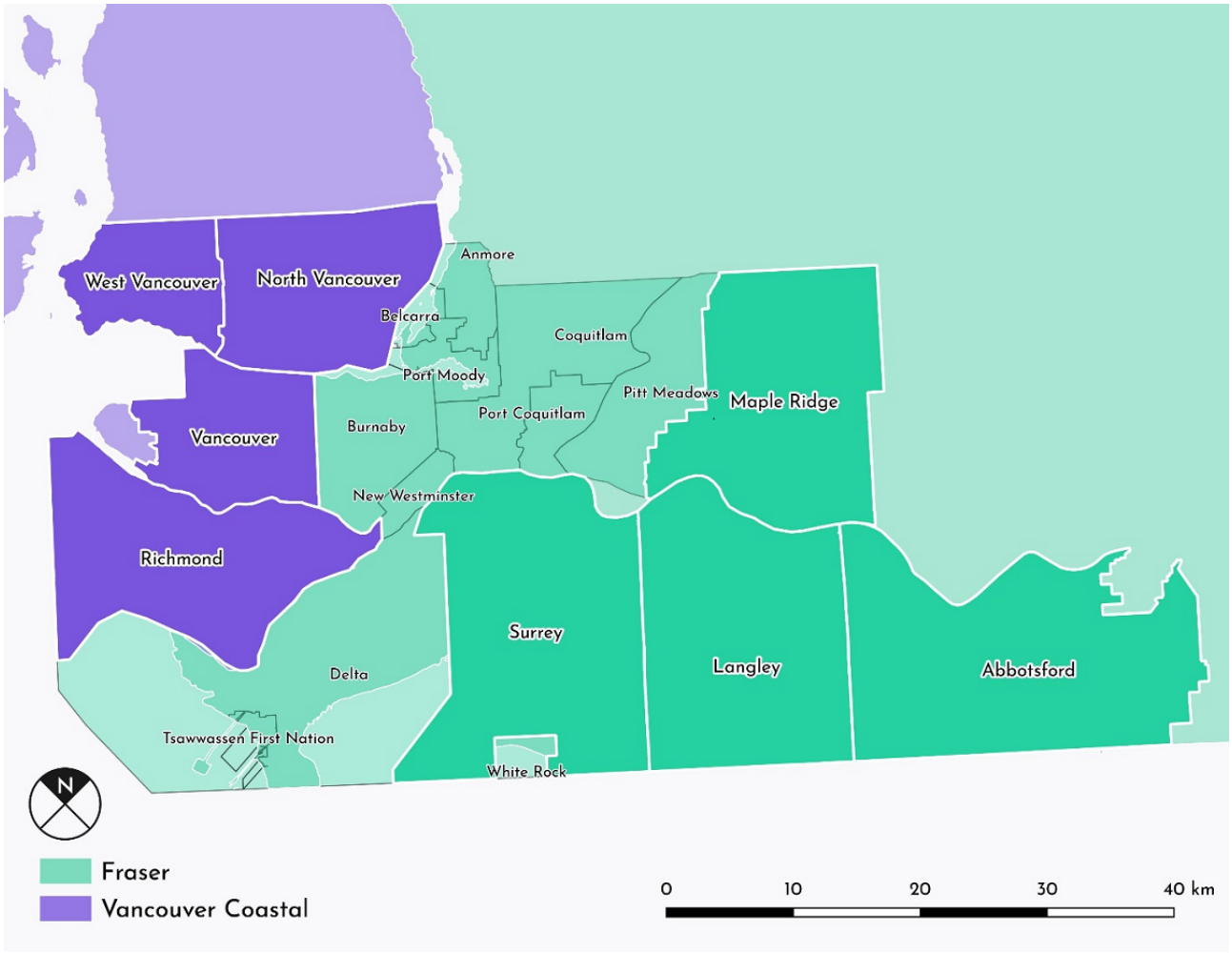
Working from the premise that medical crowdfunding campaigns may provide unique user-focused insight into local health system gaps and deficiencies, in this exploratory analysis, we examine campaigns from two administrative regions of British Columbia. These administrative areas cover a range of communities, including dense urban centers, suburban regions, and less densely populated rural areas, all geographically contiguous and with access to extensive health system infrastructure. Our aim in undertaking this analysis was to better understand what health-related needs individuals in this area are crowdfunding for, how these needs compare with the existing commentary on health system deficiencies, and the advantages and limitations of using crowdfunding campaigns to enhance or augment our understanding of perceived health system deficiencies. We further aimed for this analysis to serve as a model for other analyses using crowdfunding data to enhance understanding of health-related needs.

Methods

We used an automated Web scraper to extract campaign data from GoFundMe.com. This scraper, the Crowdfunding for Health Research Portal, began recording data from GoFundMe crowdfunding campaigns in April 2019 using the GoFundMe.com sitemap. Information for every campaign listed on the sitemap was recorded at that time, including the title, text, updates, number of donors, money requested, money pledged, number of Facebook shares, and campaigner location. All campaigns listing locations in the Fraser Health Authority (FHA) region (n=2772) and metropolitan Vancouver portion of Vancouver Coastal Health Authority (VCHA) region

(Vancouver, Richmond, North Vancouver, and West Vancouver; n=1971) were selected and recorded on a shared spreadsheet (see Figure 1). These regions are relatively densely populated geographically contiguous areas that are home to extensive health system infrastructure.

Figure 1. A map of campaigner locations.



These campaigns were then restricted to fundraising categories, most likely to be health related (emergencies, family, medical, and uncategorized), and initiated in 2018.

After campaigns were extracted from the portal, the second author reviewed the campaigns, including those raising funds to allow treatment of a medical problem or related to needs arising from ill health. All authors independently reviewed 50 of the included campaigns and met to identify categories of health-related needs motivating crowdfunding, understood as needs triggered by, exacerbated by, or otherwise related to the underlying health condition. Included campaigns were then coded by JS and MZ for the requested health-related need, and the second author recorded the underlying health problem for each campaign. The fourth author then reviewed any codes flagged as uncertain by the first 2 reviewers and audited 5% of the coded campaigns to ensure consistency. Any disagreements were discussed and resolved by these reviewers. Given the exploratory purview of this analysis, sums were generated for the coded categories to allow for descriptive statistics to be generated.

Results

We identified 423 campaigns for health-related needs between VCHA and FHA. These campaigns requested Can $8,715,806 (US $6,088,078) in funding and were pledged Can $3,477,384 (US $2,428,987) from 27,773 donors and shared 120,665 times on Facebook (see Table 1). In the FHA, campaigns were located most commonly in the cities of Surrey (47/275, 17.1%), Abbotsford (45/275, 16.4%), Langley (31/275, 11.3%), and Maple Ridge (26/275, 9.5%). In the VCHA, campaigns were divided between Vancouver (108/148, 73.0%), North Vancouver (27/148, 18.2%), Richmond (10/148, 6.8%), and West Vancouver (3/148, 2.0%; see Table 2). The most common underlying medical condition for campaign recipients was cancer, followed by traumatic injuries from collisions, and brain injury and stroke (see Table 3).

Table 1. Campaign engagement.

Category	Fraser Health Authority	Vancouver Coastal Health Authority	Total
Campaigns, n	275	148	423
Amount requested (Can $)	5,982,654 (US $4,178,944)	2,733,152	8,715,806
Money pledged (Can $)	1,939,932 (US $1,355,062)	1,537,452	3,477,384
Percentage pledged, %	32.4	56.3	39.9
Donors, n	15,826	11,947	27,773
Facebook shares, n	78,966	41,699	120,665

Table 2. Campaign location.

Location	Number of campaigns (N=423), n (%)	Population [20] (N=2,916,414), n (%)
Vancouver	108 (25.5)	672,963 (23.08)
Surrey	47 (11.1)	569,065 (19.51)
Abbotsford	45 (10.6)	151,923 (5.21)
Langley	31 (7.3)	154,867 (5.31)
North Vancouver	27 (6.4)	147,555 (5.06)
Maple Ridge	26 (6.1)	88,626 (3.00)
Burnaby	22 (5.2)	248,476 (8.52)
Chilliwack	21 (5.0)	90,931 (3.12)
Coquitlam	17 (4.0)	149,490 (5.16)
Port Coquitlam	17 (4.0)	62,844 (2.15)
New Westminster	12 (2.8)	76,799 (2.63)
Mission	10 (2.4)	41,503 (1.42)
Richmond	10 (2.4)	216,300 (7.42)
Delta	9 (2.1)	109,484 (3.75)
White Rock	6 (1.4)	21,370 (0.73)
Port Moody	4 (0.9)	35,613 (1.22)
Hope	3 (0.7)	6659 (0.23)
Pitt Meadows	3 (0.7)	19,772 (0.67)
West Vancouver	3 (0.7)	44,866 (1.54)
Agassiz	1 (0.2)	6624 (0.23)
Agassiz and Belcarra	1 (0.2)	684 (0.02)

Table 3. Underlying medical condition.

Medical condition	Fraser Health Authority (N=280), n (%)	Vancouver Coastal Health Authority (N=149), n (%)	Total (N=449), n (%)
Cancer	124 (44.3)	50 (33.6)	174 (40.6)
Trauma injuries	22 (7.9)	19 (12.8)	41 (9.6)
Brain injury/stroke	21 (7.5)	9 (6.0)	30 (7.0)
Unspecified/undiagnosed	22 (7.9)	6 (4.0)	28 (6.5)
Other	14 (5.0)	14 (9.4)	28 (6.5)
Heart disease/attack, diseases of circulatory system	14 (5.0)	4 (2.7)	18 (4.2)
Spinal disease/damage	8 (2.9)	7 (4.7)	15 (3.5)
Other neurological diseases	9 (3.2)	4 (2.7)	13 (3.0)
Kidney disease	8 (2.9)	3 (2.0)	11 (2.6)
Gastrointestinal disease/Crohn disease	5 (1.8)	6 (4.0)	11 (2.6)
Other genetic disorders	6 (2.1)	3 (2.0)	9 (2.1)
Diabetes	5 (1.8)	3 (2.0)	8 (1.9)
Gender affirmation	1 (0.4)	6 (4.0)	7 (1.6)
Amputation	3 (1.1)	4 (2.7)	7 (1.6)
Mental health and addiction treatment	2 (0.7)	5 (3.4)	7 (1.6)
Cerebral palsy/muscular dystrophy	5 (1.8)	1 (0.7)	6 (1.4)
Lyme disease	5 (1.8)	1 (0.7)	6 (1.4)
Burn injuries	5 (1.8)	0 (0.0)	5 (1.2)
Multiple sclerosis	1 (0.4)	4 (2.7)	5 (1.2)

After discussion, 13 categories of health-related needs motivating crowdfunding were identified. Campaigners were motivated by seeking funding for a variety of issues related to the recipient's health needs (see Table 4). Of the 685 motivations recorded, including multiple motivations in several instances, by far, the most common was seeking financial support for wages lost because of illness (232/684, 33.9%). These included instances of income lost because of taking time off of paid employment because of the effects of illness, needing to take time off of paid employment for treatment, and needing to relocate for treatment. Such needs were often prompted by recipients reaching the limits in unemployment insurance coverage, as for the campaigner who wrote that "employment insurance does not cover much and only lasts for 10 weeks."

Table 4. Motivation.

Motivation category	Fraser Health Authority (N=444), n (%)	Vancouver Coastal Health Authority (N=231), n (%)	Total (N=675), n (%)
Lost wages	153 (34.5)	79 (34.2)	232 (34.4)
Local travel expenses	35 (7.9)	14 (6.1)	49 (7.3)
Out-of-province travel	29 (6.5)	20 (8.7)	49 (7.3)
Medical equipment and supplies	43 (9.7)	22 (9.5)	65 (9.6)
Home accessibility improvements	16 (3.6)	3 (1.3)	19 (2.8)
Lack of travel insurance	0 (0.0)	2 (0.9)	2 (0.3)
Private care because of wait times	3 (0.7)	4 (1.7)	7 (1.0)
Private care because of quality of service	3 (0.7)	3 (1.3)	6 (0.9)
Medications	28 (6.3)	15 (6.5)	43 (6.4)
Caregiver expenses	11 (2.5)	1 (0.4)	12 (1.8)
Complementary, alternative, or unproven treatments	34 (7.7)	22 (9.5)	56 (8.3)
Elective or uninsured direct medical expenses	36 (8.1)	22 (9.5)	58 (8.6)
Unspecified or other direct medical costs	53 (11.9)	24 (10.4)	77 (11.4)

Other common motivations for crowdfunding for health-related expenses included elective or uninsured direct medical expenses (68/685, 9.9%). The treatments sought included diagnostic testing, fertility treatments, physiotherapy, and dental treatment, among others. Some campaigns (65/685, 9.5%) sought help with purchasing medical equipment and supplies. This equipment was typically intended for home use after discharge from hospital. Other campaigns (56/685, 8.2%) sought to fund complementary, alternative, or unproven treatments including experimental interventions. These requests ranged from complementary cancer treatments such as acupuncture to naturopathic treatment for Lyme disease and unproven stem cell interventions abroad. Seeking financial support to cover travel-related costs was another common motivation for crowdfunding, including in-province (49/685, 7.2%) and out-of-province (49/685, 7.2%) travel. Those traveling in-province typically needed help with gas expenses and parking for medical appointments or relocation to be closer to a medical center. Out-of-province travel included help relocating within Canada or, more commonly, travel abroad to privately purchase medical care. Help paying for medications was requested in 6.3% (43/685) campaigns. These were typically prescription medications, as in the case where “some of my medical issues require prescriptions that are not covered which means that I can't get them.” Unspecified health-related expenses were the second largest category (77/685, 11.2%). All other motivation categories were under 5% of all instances.

Discussion

Principal Findings

Our exploratory findings show the scope of crowdfunding in one metropolitan region in Canada. Despite universal public insurance coverage for Canadian residents, at least 423 individuals living in the areas covered by FHA and VCHA initiated crowdfunding campaigns in 2018, raising nearly Can $3.5 million (US $2,541,886). These campaigns impacted a wide network, as they received nearly 28,000 pledges of donations and were shared on Facebook over 120,000 times. Given that these campaigns may have been shared on other social media platforms and each share is viewed by multiple individuals, the impact of these campaigns is considerable.

As has been reported elsewhere, seeking funds to assist with addressing the costs of treating and managing cancer is by far the most common underlying health issue motivating crowdfunding by Canadians [21]. Other medical conditions linked to the crowdfunding campaigns reviewed in this analysis are less discussed in this literature, including trauma injuries because of collisions, brain injury and stroke, heart disease, and spinal injuries. This may reflect specific health needs in the region analyzed or broader crowdfunding trends across Canada.

Our findings demonstrate that out-of-pocket expenses incurred by people who were accessing care provided by the publicly funded health system were by far the most commonly cited reasons for using medical crowdfunding. Lost wages because of illness and accessing treatment was by far the most common motivation, appearing in 54.8% (232/423) campaigns. This may indicate insufficient access to or other limitations in employment benefits for residents. Indirect expenses such as in-province and out-of-province travel to access care (98/423, 23.2%), purchasing medical equipment and supplies (65/423, 15.4%), and making home accessibility improvements were also common (19/423, 4.5%). Although some people sought funds to cover the cost of privately-paid-for medical care, the overwhelming majority of campaigns was effectively illustrating the limits of the support available to those accessing the public system.

The extant literature on barriers to Canadians accessing medical care focuses on wait times for care. And there is some concern that these wait times drive Canadians to privately purchase care domestically or abroad. Meanwhile, this motivation was rarely found in the campaigns reviewed in this analysis, appearing in only 1.0% (7/100) instances. Similarly, accessing care privately because of the perceived (low) quality of care in the publicly funded system appeared only 0.9% (6/700) times. When the direct provision of care motivated crowdfunding, it was generally not to access private care more quickly but to afford elective or other uninsured forms of care (68/423, 16.1%), complementary or alternative care not covered by insurance (56/423, 13.2%), and medication not covered by insurance (43/423, 10.2%).

These findings demonstrate that campaigns exhibit spatial differences regarding crowdfunding requests between those living in the FHA and VCHA areas. Although both FHA and VCHA campaigns listed cancer as the most common underlying health condition motivating the campaign, FHA and VCHA contained 44.3% and 33.6% of cancer-related campaigns, respectively. Similarly, heart and circulatory diseases and kidney disease were much more common in FHA. Conversely, 7.9% of FHA campaigns involved accident trauma injuries compared with 12.3% for VCHA, and 4% of VCHA campaigns were for gender affirmation treatments compared with 0.4% for FHA. Motivations for engaging in crowdfunding were generally consistent across this region, although home accessibility improvements were nearly three times more common in FHA and private care because of wait times was more than twice as common in VCHA. Usage of crowdfunding varied greatly across metro Vancouver and the Fraser Valley as well, with higher usage rates compared with the underlying population for Abbotsford and Maple Ridge and lower usage rates for Richmond and Surrey. These differences merit continued exploration, as they could reflect spatially specific difficulties in meeting health-related needs.

The most striking difference between these two regions was in terms of success in meeting crowdfunding goals. FHA campaigns were pledged 32.4% of their requests compared with VCHA’s 56.3%. Evidence is emerging that crowdfunding campaigns may increase inequities by linking fundraising success to socioeconomic advantages and disadvantaging marginalized communities [22,23]. Average total annual income per person is Can $93,808 (US $65,417; Vancouver) and Can $83,850 (US $58,472; Richmond) in the VCHA vs Can $84,023 (US $58,593; Fraser East), Can $90,386 (US $63,031; Fraser North), and Can $97,301 (US $67,853; Fraser South) [24]. Statistics Canada distinguishes these two health authorities as falling into distinctly separate peer groups, with FHA having greater rates of visible minorities compared with VCHA (very

high vs high) and Aboriginal residents (low vs very low) [25]. Life expectancies in VCHA and FHA are 84.4 and 82.8, respectively, and the infant mortality rates for VCHA and FHA are 3.0 and 3.3, respectively [26]. These socioeconomic and health inequality differences may be significant factors driving the differences in meeting campaign goals between those living in the FHA and VCHA regions. More generally, these findings raise questions as to whether the socioeconomic differences between regions are associated with differences in crowdfunding campaign success. However, this falls outside the scope of the current analysis and thus raises important implications for future research directions.

These findings can be used to support some existing arguments regarding the impact of insurance gaps and indirect medical expenses as pressing needs facing Canadians. Given current debates over medical prescription costs and PharmaCare expansion, this analysis demonstrates that such expenses are pushing some Canadians to turn to others for financial support to meet drug costs [27]. The prevalence of lost wages and travel expenses in these crowdfunding campaigns also show why indirect medical expenses must be part of the conversation of the social costs of ill health and gaps in social insurance. For example, this reflects public debates in British Columbia, where fees for parking costs at hospitals have been criticized as exploitative and unfair [28]. In total, 56 (13.2%) campaigns sought complementary and alternative medical interventions that, in some cases, are not evidence based, potentially risky to patients, and contain misinformation that is spread via these crowdfunding campaigns [29]. This finding can help inform debates about public funding for complementary and alternative treatments and confusion about their relationship to traditional, evidence-based treatment [30].

It is notable that some underlying health needs were largely not represented in these campaigns despite a prominent place in local discussion of the health system. Much recent discussion has taken place in metropolitan Vancouver around the opioid crisis, mental health and addiction issues, and the lack of adequate access to treatment to address these health needs [31]. However, these conditions were represented in only 1.6% (7/438) of the campaigns reviewed. This suggests that we should use campaigns to complement ongoing local dialog about health care but not replace other sources. Moreover, some of the most pressing health system priorities are a clear reflection of the socioeconomic gradient in communities. Noting that people were not crowdfunding for this care may add support to the literature that suggests that those who crowdfund are more educated and have higher socioeconomic status than the community at large.

Limitations

This analysis has two main limitations, given the data used. First, crowdfunding data are self-reported and may be inaccurate. For example, the location of the campaign reflects the location of the campaign maker, who may be different from and in a different location than the campaign recipient. Campaigns were carefully reviewed to exclude cases where the recipient was clearly outside of the geographic area of interest. Second, because crowdfunding campaigns can be deleted, and data were collected in early 2019, many campaigns from the VHA and VCHA regions were likely not included in our findings. Thus, the findings here likely represent an undercount of the total crowdfunding activity in the FHA and VCHA.

Conclusions

This analysis demonstrates the potential of crowdfunding data to present timely and context-specific user-created insights into the perceived health-related financial needs of some Canadians. Such insights can complement those generated by other sources, such as administrative and census data and even media and public discussions about health system reforms. New crowdfunding campaigns are being generated on an ongoing basis campaign-based data can be used to identify trends in perceived health-related financial needs as they develop as well as service gaps. These data are also context specific and thus can help to inform policy makers, patient advocates, health workers, and other stakeholders' perceived health system deficiencies that lead to financial burdens for residents of Canada. Importantly, not all of the health-related needs described in these campaigns should be interpreted as identifying health and social system gaps that must be addressed. That said, a better understanding of trends around practices that fall well outside the scope of the Canada Health Act, such as accessing unproven medical interventions and traveling abroad for care, are also useful for health system stakeholders. This is because practices such as these have been shown to potentially introduce health system burdens and create ethical challenges for physicians [32].

We encourage the development of new resources to harness the power of crowdfunding data as a supplementary source of information for Canadian health system stakeholders. Although the findings of this study demonstrate the potential of these data to shed additional light on perceived health system deficiencies, they also show that context is of great importance in using these data. Despite their geographic proximity, the crowdfunding campaigns from these two health regions showed noteworthy differences between the percentages of requested funding that received. This finding lends support to existing questions about what factors determine crowdfunding success and whether this reflects larger social inequities. Finally, more research on crowdfunding campaigns across Canada, including those in rural settings, will allow regional differences in crowdfunding motivations and success to be identified.

Authors' Contributions

All authors reviewed the data, identified motivation themes, and edited the manuscript. JS wrote the manuscript. MZ cleaned and coded all data.

Conflicts of Interest

None declared.

References

1. Martin D, Miller A, Quesnel-Vallée A, Caron N, Vissandjée B, Marchildon G. Canada's universal health-care system: achieving its potential. Lancet 2018 Apr 28;391(10131):1718-1735. [doi: 10.1016/S0140-6736(18)30181-8] [Medline: 29483027]
2. Valle VM. An assessment of Canada's healthcare system weighing achievements and challenges. Norteamérica 2016 Dec 20;11(2):193-220 [FREE Full text] [doi: 10.20999/nam.2016.b008]
3. Simpson C, Walker DM, Drummond D, Sinclair D, Wilson R. The Conversation. 2017 Sep 25. How Healthy is the Canadian Health-Care System? URL: http://theconversation.com/how-healthy-is-the-canadian-health-care-system-82674 [accessed 2020-03-13]
4. Clark J. Statistics Canada. 2016 Dec 8. Difficulty Accessing Health Care Services in Canada URL: https://www150.statcan.gc.ca/n1/pub/82-624-x/2016001/article/14683-eng.htm [accessed 2020-03-13]
5. Canadian Institute for Health Information. 2019 Mar 28. Wait Times for Priority Procedures in Canada URL: https://www.cihi.ca/en/wait-times-for-priority-procedures-in-canada [accessed 2020-03-13]
6. Canadian Institute for Health Information. CIHI Wait Times. URL: http://waittimes.cihi.ca/ [accessed 2020-03-13]
7. Steinbrook R. Private health care in Canada. N Engl J Med 2006 Apr 20;354(16):1661-1664. [doi: 10.1056/NEJMp068064] [Medline: 16625005]
8. Palley HA, Pomey MP, Forest PG. Examining private and public provision in Canada's provincial health care systems: comparing Ontario and Quebec. Int Polit Sci Rev 2011;32(1):79-94 [FREE Full text] [doi: 10.1177/0192512110385291]
9. Johnston R, Crooks V, Snyder J. 'I didn't even know what I was looking for': A qualitative study of the decision-making processes of Canadian medical tourists. Global Health 2012 Jul 7;8:23 [FREE Full text] [doi: 10.1186/1744-8603-8-23] [Medline: 22769723]
10. Marchildon G. Canada: Health system review. Health Syst Transit 2013;15(1):1-179 [FREE Full text] [Medline: 23628429]
11. Law MR, Daw JR, Cheng L, Morgan SG. Growth in private payments for health care by Canadian households. Health Policy 2013 May;110(2-3):141-146. [doi: 10.1016/j.healthpol.2013.01.014] [Medline: 23419418]
12. Longo CJ, Deber R, Fitch M, Williams AP, D'Souza D. An examination of cancer patients' monthly 'out-of-pocket' costs in Ontario, Canada. Eur J Cancer Care (Engl) 2007 Nov;16(6):500-507. [doi: 10.1111/j.1365-2354.2007.00783.x] [Medline: 17944764]
13. Obembe AO, Goldsmith CH, Simpson LA, Sakakibara BM, Eng JJ. Support service utilization and out-of-pocket payments for health services in a population-based sample of adults with neurological conditions. PLoS One 2018;13(2):e0192911 [FREE Full text] [doi: 10.1371/journal.pone.0192911] [Medline: 29474391]
14. Longo CJ, Bereza BG. A comparative analysis of monthly out-of-pocket costs for patients with breast cancer as compared with other common cancers in Ontario, Canada. Curr Oncol 2011 Jan;18(1):e1-e8 [FREE Full text] [doi: 10.3747/co.v18i1.681] [Medline: 21331267]
15. Coyte PC, McKeever P. Home care in Canada: passing the buck. Can J Nurs Res 2000;33(2):11-25 [FREE Full text]
16. Daw J, Morgan S. Stitching the gaps in the Canadian public drug coverage patchwork?: a review of provincial pharmacare policy changes from 2000 to 2010. Health Policy 2012 Jan;104(1):19-26 [FREE Full text] [doi: 10.1016/j.healthpol.2011.08.015] [Medline: 21978939]
17. Lunden I. TechCrunch. 2017 Jan 11. GoFundMe Acquires CrowdRise to Expand to Fundraising for Charities URL: https://techcrunch.com/2017/01/10/gofundme-buys-crowdrise-to-expand-to-fundraising-for-charities/?guccounter=1 [accessed 2020-03-13]
18. Harris A. Fast Company. 2017 Feb 13. How Crowdfunding Platform GoFundMe Has Created A $3 Billion Digital Safety Net URL: https://www.fastcompany.com/3067472/how-crowdfunding-platform-gofundme-has-created-a-3-billion-digital [accessed 2020-03-13]
19. Heller N. The New Yorker. 2019 Jun 24. The Hidden Cost of GoFundMe Health Care URL: https://www.newyorker.com/magazine/2019/07/01/the-perverse-logic-of-gofundme-health-care [accessed 2020-03-13]
20. Ministry of Citizens Services. Province of British Columbia - Government of BC. Population Estimates URL: https://www2.gov.bc.ca/gov/content/data/statistics/people-population-community/population/population-estimates [accessed 2020-03-13]
21. Snyder J, Crooks VA, Mathers A, Chow-White P. Appealing to the crowd: ethical justifications in Canadian medical crowdfunding campaigns. J Med Ethics 2017 Jun;43(6):364-367 [FREE Full text] [doi: 10.1136/medethics-2016-103933] [Medline: 28137998]
22. van Duynhoven A, Lee A, Michel R, Snyder J, Crooks V, Chow-White P, et al. Spatially exploring the intersection of socioeconomic status and Canadian cancer-related medical crowdfunding campaigns. BMJ Open 2019 Jun 20;9(6):e026365 [FREE Full text] [doi: 10.1136/bmjopen-2018-026365] [Medline: 31227531]

23. Kenworthy N, Dong Z, Montgomery A, Fuller E, Berliner L. A cross-sectional study of social inequities in medical crowdfunding campaigns in the United States. PLoS One 2020;15(3):e0229760 [FREE Full text] [doi: 10.1371/journal.pone.0229760] [Medline: 32134984]
24. Statistics Canada. 2017 Dec. Census Profile, 2016 Census URL: https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=HR&Code1=5931&Geo2=HR&Code2=5921&SearchText=Fraser%20East%20Health%20Service%20Delivery%20Area&SearchType=Begins&SearchPR=01&B1=All&GeoLevel=PR&GeoCode=5921&TABID=1&type=0 [accessed 2020-03-13]
25. Statistics Canada. 2017 Dec. Table 9 Summary Table of Peer Groups and Principal Characteristics URL: https://www150.statcan.gc.ca/n1/pub/82-402-x/2015001/regions/tbl/tbl9-eng.htm [accessed 2020-03-13]
26. Office of the Provincial Health Officer. Province of British Columbia - Government of BC. 2019 Jan. Taking the Pulse of the Population: An Update on the Health of British Columbians URL: https://www2.gov.bc.ca/gov/content/health/about-bc-s-health-care-system/office-of-the-provincial-health-officer/reports-publications/annual-reports [accessed 2020-03-13]
27. Flood C, Thomas B, Moten A, Fafard P. Institute for Research on Public Policy. 2018 Sep 12. Universal Pharmacare and Federalism: Policy Options for Canada URL: https://irpp.org/research-studies/universal-pharmacare-and-federalism-policy-options-for-canada/ [accessed 2020-03-13]
28. Robinson M. Vancouver Sun. 2019 Jan 22. Time to End 'Exploitative' Hospital Parking Fees in BC: Non-Profit URL: https://vancouversun.com/news/local-news/time-to-end-exploitative-hospital-parking-fees-in-b-c-non-profit [accessed 2020-03-13]
29. Vox F, Folkers KM, Turi A, Caplan AL. Medical crowdfunding for scientifically unsupported or potentially dangerous treatments. J Am Med Assoc 2018 Oct 23;320(16):1705-1706 [FREE Full text] [doi: 10.1001/jama.2018.10264] [Medline: 30357284]
30. Kirkey S. National Post. 2019 Nov 5. Criticism for Ontario City's Decision to Grant $45K Toward Naturopathic Clinic URL: https://nationalpost.com/news/confused-ontario-city-spends-45k-toward-naturopathic-clinic-because-it-has-a-shortage-of-doctors [accessed 2020-03-13]
31. Johnston G. The kids are all white: examining race and representation in news media coverage of opioid overdose deaths in Canada. Sociol Inq 2020 Feb;90(1):123-146 [FREE Full text] [doi: 10.1111/soin.12269]
32. Crooks V, Snyder J. Medical tourism: what Canadian family physicians need to know. Can Fam Physician 2011 May;57(5):527-9, e151 [FREE Full text] [Medline: 21571709]

Abbreviations

FHA: Fraser Health Authority
VCHA: Vancouver Coastal Health Authority

Edited by G Eysenbach; submitted 08.11.19; peer-reviewed by M Lisboa, J Zhang; comments to author 13.02.20; revised version received 15.02.20; accepted 26.02.20; published 22.05.20.

Please cite as:
Snyder J, Zenone M, Crooks V, Schuurman N
What Medical Crowdfunding Campaigns Can Tell Us About Local Health System Gaps and Deficiencies: Exploratory Analysis of British Columbia, Canada
J Med Internet Res 2020;22(5):e16982
URL: http://www.jmir.org/2020/5/e16982/
doi:10.2196/16982
PMID:32348269

Original Paper

Exploring the Vast Choice of Question Prompt Lists Available to Health Consumers via Google: Environmental Scan

Marguerite Clare Tracy[1*], MBBS (Hons), MPH, FRACGP; Heather L Shepherd[1,2*], BA (Hons), Dip HE, PhD; Pinika Patel[1*], BMedSci, MPH; Lyndal Jane Trevena[1*], MBBS (Hons), MPhilPH, PhD

[1]Ask, Share, Know: Rapid Evidence for General Practice Decisions Centre for Research Excellence, School of Public Health, The University of Sydney, NSW, Australia

[2]Centre for Medical Psychology and Evidence-based Decision-making (CeMPED), The University of Sydney, NSW, Australia

[*]all authors contributed equally

Corresponding Author:
Marguerite Clare Tracy, MBBS (Hons), MPH, FRACGP
Ask, Share, Know: Rapid Evidence for General Practice Decisions Centre for Research Excellence
School of Public Health
The University of Sydney
Room 323A, Edward Ford Building A27
NSW, 2006
Australia
Phone: 61 293512942
Email: marguerite.tracy@sydney.edu.au

Abstract

Background: There is increasing interest in shared decision making (SDM) in Australia. Question prompt lists (QPLs) support question asking by patients, a key part of SDM. QPLs have been studied in a variety of settings, and increasingly the internet provides a source of suggested questions for patients. Environmental scans have been shown to be useful in assessing the availability and quality of online SDM tools.

Objective: This study aimed to assess the number and readability of QPLs available to users via Google.com.au.

Methods: Our environmental scan used search terms derived from literature and reputable websites to search for QPLs available via Google.com.au. Following removal of duplicates from the 4000 URLs and 22 reputable sites, inclusion and exclusion criteria were applied to create a list of unique QPLs. A sample of 20 QPLs was further assessed for list length, proxy measures of quality such as a date of review, and evidence of doctor endorsement. Readability of the sample QPL instructions and QPLs themselves was assessed using Flesch Reading Ease and Flesch-Kincaid Grade Level scores.

Results: Our environmental scan identified 173 unique QPLs available to users. Lists ranged in length from 1 question to >200 questions. Of our sample, 50% (10/20) had a listed date of creation or update, and 60% (12/20) had evidence of authorship or source. Flesch-Kincaid Grade Level scores for instructions were higher than for the QPLs (grades 10.3 and 7.7, respectively). There was over a 1 grade difference between QPLs from reputable sites compared with other sites (grades 4.2 and 5.4, respectively).

Conclusions: People seeking questions to ask their doctor using Google.com.au encounter a vast number of question lists that they can use to prepare for consultations with their doctors. Markers of the quality or usefulness of various types of online QPLs, either surrogate or direct, have not yet been established, which makes it difficult to assess the value of the abundance of lists. Doctor endorsement of question asking has previously been shown to be an important factor in the effectiveness of QPLs, but information regarding this is not readily available online. Whether these diverse QPLs are endorsed by medical practitioners warrants further investigation.

(J Med Internet Res 2020;22(5):e17002) doi:10.2196/17002

KEYWORDS

question prompt list; shared decision making; environmental scan

Introduction

The role of shared decision making (SDM) as a part of patient-centered care in clinical consultations is being increasingly recognized as having positive outcomes for patients [1]. Internationally, there are efforts to make SDM a part of routine health care [2]. The National Safety and Quality Health Service Standards (Second Edition) developed by the Australian Commission on Safety and Quality in Health Care include a statement that "Integral to the process is encouraging patients to be more involved and ask their doctor more questions during consultations" [3]. To facilitate patient question asking as a part of SDM, question prompt list (QPL) tools have been developed, and some have been evaluated and published in the peer-reviewed literature, some of which are available via the internet [4,5].

Increasingly, people are turning to the internet for health information [6]. Search engines are the predominant tool used by people to search for health information online [7]. In 2018, 78% of Australian adults reported using the internet to find health-related information [8]. Online QPLs are being used, and those using online question lists have been shown to prefer questions that specifically support SDM [9]. Physician endorsement has been found to be key to the successful implementation of QPLs into practice [4]. Similarly, a dominant factor for patient online health information seeking on the doctor-patient relationship is the doctor's willingness to discuss the information [10].

With an increasing focus on question asking in consultations, many websites include lists of questions for someone to take to their doctor as an additional resource [11-13]. While there has been some research into the implementation of QPLs into practice, there has not been an assessment of the prevalence of such lists available online [5]. Given that online material can be created and hosted by anyone and there is no regulation of the quality of information or available tools, a study of these QPLs is warranted [14]. Assessing the readability of QPL resources as well as their prevalence is important in addressing usefulness.

Adequate literacy and health literacy of patients are important in the use of tools to support question asking and SDM [15]. In their systematic review published in 2012, Sørensen et al [16] identified several competencies of health literacy around having the ability to access, understand, and use information to make decisions about health. Supporting people with lower health literacy by making such tools readable and accessible is one way to increase access to SDM [15]. Several standards exist for the readability of patient information and other SDM tools such as decision aids, to increase the accessibility of materials for target audiences [17].

While standards exist for assessing the quality of patient information resources, such as the adaption of DISCERN for internet information [18,19], there is ongoing difficulty in assessing the quality of online medical information and resources as well as how users perceive and use that information [20-23]. In addition, there are factors other than the information itself that influence information preferences, such as domain bias [24] and webpage design [21]. For example, the extension ".com" is used for commercial sites, while ".edu," ".gov," and others denote non-commercial or government sites. Domain extensions have some bearing on how users view the information provided and their trust in the source [24,25].

Environmental scan processes have previously been used to assess available online decision aids and risk calculators [26,27]. They allow a real-time snapshot of the availability of online resources available to users. The aim of this study was to conduct an environmental scan to describe and assess the number and readability of QPLs readily available to health consumers in the online environment.

Methods

Overview

We used a previously published methodology to conduct an environmental scan to search Google.com.au for question lists relating to patient-doctor clinical interactions [27]. Google.com.au was chosen as it is the most frequently used search engine in Australia (94.11% as of Dec 2018) [28]. To assess the number of question lists available to health care consumers online, 2 stages were needed. A third stage involved the assessment of the readability of a sample of QPLs and their instructions.

Stage 1: Choice of Reputable Websites and Search Term Development

The authors have expertise and backgrounds in nursing and medicine; knowledge of organizations and websites used by health professionals and consumers was used to identify a range of reputable organizations' websites to reflect the clinical areas in which QPLs have been studied [4]. The aim was to ensure that selected sites included both disease-specific lists (eg, cancer websites, parent information about their child's attention deficit hyperactivity disorder) as well as sites with more generic lists, such as those with consumer health information. We also aimed to ensure a mix of local (Australian) and international organizations with patient-focused information. The final list of 22 URLs for these organizations was decided by consensus between the authors (Table 1). These websites were accessed via the URLs to confirm that they referred to, or included, QPLs.

Using previous systematic reviews of QPLs and citation snowballing, we found 11 terms in the published literature that have been used to describe patient question lists [4,29]. In addition, the reputable organization list websites were accessed to find the language and terms used to describe QPLs on these sites; there were a further 9 terms found. Using these 2 sources, we had a total of 20 search terms for use in Stage 2 of the scan (Table 2).

Table 1. URLs of the 22 selected reputable websites.

Site or organization type, organization	URL
Consumer-directed organizations	
Healthdirect Australia Ltd.	healthdirect.gov.au
Consumers Health Forum	chf.org.au
Information about medicine or prescribing	
National Prescribing Service – Choosing Wisely	choosingwisely.org.au/home
National Prescribing Service	nps.org.au
Government entities	
Department of Health	health.gov.au
Health Canada	canada.ca/en/health-canada
NHS[a] United Kingdom	nhs.uk
Therapeutic Goods Administration	tga.gov.au
NHS Networks (Ask 3 questions)	personcentredcare.health.org.uk/resources/ask-3-questions-materials
Institute for Healthcare Improvement (Ask Me 3)	ihi.org/resources/Pages/Tools/Ask-Me-3-Good-Questions-for-Your-Good-Health.aspx
Disease-specific organizations	
Cancer Council Australia	cancer.org.au
Cancer Australia	canceraustralia.gov.au
Raising Children Network	raisingchildren.net.au
ADHD[b] Australia	adhdaustralia.org.au
Breast Cancer Network Australia	bcna.org.au
Family Planning NSW[c]	fpnsw.org.au
Pediatric organizations	
Royal Children's Hospital (RCH) Melbourne	rch.org.au
Sydney Children's Hospital Network	schn.health.nsw.gov.au
HealthyChildren.org (American Academy of Pediatrics)	healthychildren.org
Quality and safety organizations	
Australian Commission for Safety and Quality in Health Care	safetyandquality.gov.au
Agency for Healthcare Research and Quality	ahrq.gov
Wiser Healthcare	wiserhealthcare.org.au

[a]NHS: National Health Service.

[b]ADHD: attention deficit hyperactivity disorder.

[c]NSW: New South Wales.

Table 2. Search terms derived from reputable sites and literature.

Search term	Source
Ask health professional questions	Wiser Healthcare
Ask for information	NPS[a] MedicineWise
Questions to ask your doctor	Choosing Wisely, Cancer Council, Healthdirect
Questions to ask	Family Planning NSW[b]
Asking questions	Choosing wisely
Ask your health professional	NPS MedicineWise
Patient ask questions	Scottish Health Council (Ask Me 3 search)
Patients ask provider	Ask Me 3
Asking (these 3) questions during appointment	Ask 3 questions
Question prompt list	Literature
Question prompt sheet	Literature
Patient question prompt list	Literature
Patient question prompt sheet	Literature
Question sheet	Literature
Question list	Literature
Patient question aid	Literature
Shared decision-making tool	Literature
Patient agenda form	Literature
Patient agenda list	Literature
Patient question asking support tool	Literature

[a]NPS: National Prescribing Service.

[b]NSW: New South Wales.

Stage 2: Search Strategy

A systematic search of Google.com.au using each of the 20 search terms was then conducted by 2 independent researchers (MT and PP), one term at a time after clearing the browser cache. The first 100 URL results for each term searched were downloaded to Excel spreadsheets and included in the first round. Users of Google have a strong bias to the order of results presented by Google [30], and few users look for a result beyond the first results page [31]. There are only 10 results displayed on the results page with Google's default settings. We aimed to assess the breadth of lists available with our search terms, which was the reason for assessing the first 100 results for each term.

Inclusion Criteria

Websites, accessed via search URLs, included in the evaluation needed to meet the following criteria: provide a list of questions; question lists were described as for use by patients, carers, or parents in medical consultations (general practitioner or specialist medical consultations); lists were freely accessible (without registration or requiring payment); lists were written in English; and lists were visible as part of the website, not requiring downloading to be viewed, such as video files.

Exclusion Criteria

Websites or lists were excluded if the question list stated it was for doctor's use; the use of the question list required additional supporting software, such as third-party document viewers; the website required registration or incurred a cost to access the question list; the list was on a sponsored site, such as a paid site that appears before the search results; the list required downloading to be able to be viewed, such as video files; the list was provided in an academic paper, unless the journal was aimed at consumers and the list was visible; and questions list was not focused on general practice and specialist consultations (eg, counselling for entry into clinical trials). URLs blocked by the university security software system were also not included as they were deemed a potential threat.

The inclusion and exclusion criteria were then applied to the data by opening each of the URL links, again by MT and PP. Discrepancies regarding inclusion were resolved using a third reviewer (HS). Where there were duplications of lists, such as where the question list was identified on the website as sourced from another site or the webpage linked to a list already included in the review, these sites were then excluded to ascertain the final number of unique lists.

Further data about the URLs, websites, and lists were also collected. All URLs were assessed for the URL domain extension (ie, .edu, .com, .org). A sample of 20 lists was collated

to consist of 10 URLs from the reputable organizations and, for comparison, a further 10 lists from other sites were selected using random number generation and matching to URLs from the Excel spreadsheet. The sample lists were assessed for the date of creation, review, or update, if available (copyright date for the website was not considered to be an indication that the website material had been reviewed); any evidence of review of the page, such as the name of a writer or reviewer; and an assessment of the number of questions in the lists.

Stage 3: Readability

To assess the readability of the QPLs and their instructions for use, we utilized the Flesch Reading Ease (FRE) and Flesch-Kincaid Grade Level (FKGL) scores. The FRE is one measure of the complexity of a piece of text and is a score between 0-100 calculated using average sentence length and the average number of syllables of words in the text [32]. The higher the FRE score, the easier the text is to read. The FKGL uses the same data with different weightings with a result that equates to the number of years of schooling (in the United States) required to read the text [33]. We applied these formulae to the sample lists. Each of the 20 URLs were accessed again, and both the online instructions for use (or the available preamble to the question list) and then the list of question(s) were copied into separate documents. The readability tool was then applied to the copied texts to calculate FKGL and FRE scores (Table 3).

Table 3. Description of reading ease scores, reproduced from Flesch [32].

Flesch Readability Scores (reading ease), points	Description of style
90-100	Very easy
80-89	Easy
70-79	Fairly easy
60-69	Standard
50-59	Fairly difficult
30-49	Difficult
0-29	Very confusing

Results

Following removal of duplicates from the 4000 URLs and 22 reputable sites and the application of the inclusion and exclusion criteria, there were a total of 235 lists. A further review of websites revealed 62 instances of list duplication (eg, links to reputable lists and multiple uses of a list within an organization). Using our search method, 173 unique lists were identified. See Figure 1 for the study diagram of the search. There were 15 websites that had lists used, referred to, or had links to them from other websites; 9 of these were from our reputable website list, which accounted for 46 links.

We noted a wide range in the number of questions in resources, from a single question to over 200 questions in a single resource. The most common URL domain extension was .org (115/235, 48.9%), followed by .com (63/235, 26.8%), .gov (24/235, 10.2%), .edu (11/235, 4.7%), and country code extensions such as .uk and .ca (18/235, 7.7%). More detailed analysis of the sample of 20 lists found that the number of questions ranged from 3 to 56 (mean 20.5 questions, mode 3 questions, median 18 questions; Table 4). Half (10/20, 50%) of the sample lists had a date of creation or review, with the range being 2002 to the date of the preliminary review, October 29, 2018. Evidence of authorship of the QPL was available for 80% of the reputable sample sites and 40% of the other sample sites.

Table 4 also shows the readability data scores for the 20 sample lists and their online instructions. The FRE scores were higher for lists from the reputable sites compared to other sites and for question lists compared to instructions. Similarly, the grade level required to read question lists was lower than for the instructions, with reputable sites also requiring lower grade levels for readability of instructions than for other sites.

The readability was found to be easier on the reputable websites compared to the other websites, with scores showing content ranged from “very easy” to “fairly easy” for question lists and from “easy” to “difficult” for the instructions.

Figure 1. Search strategy and results.

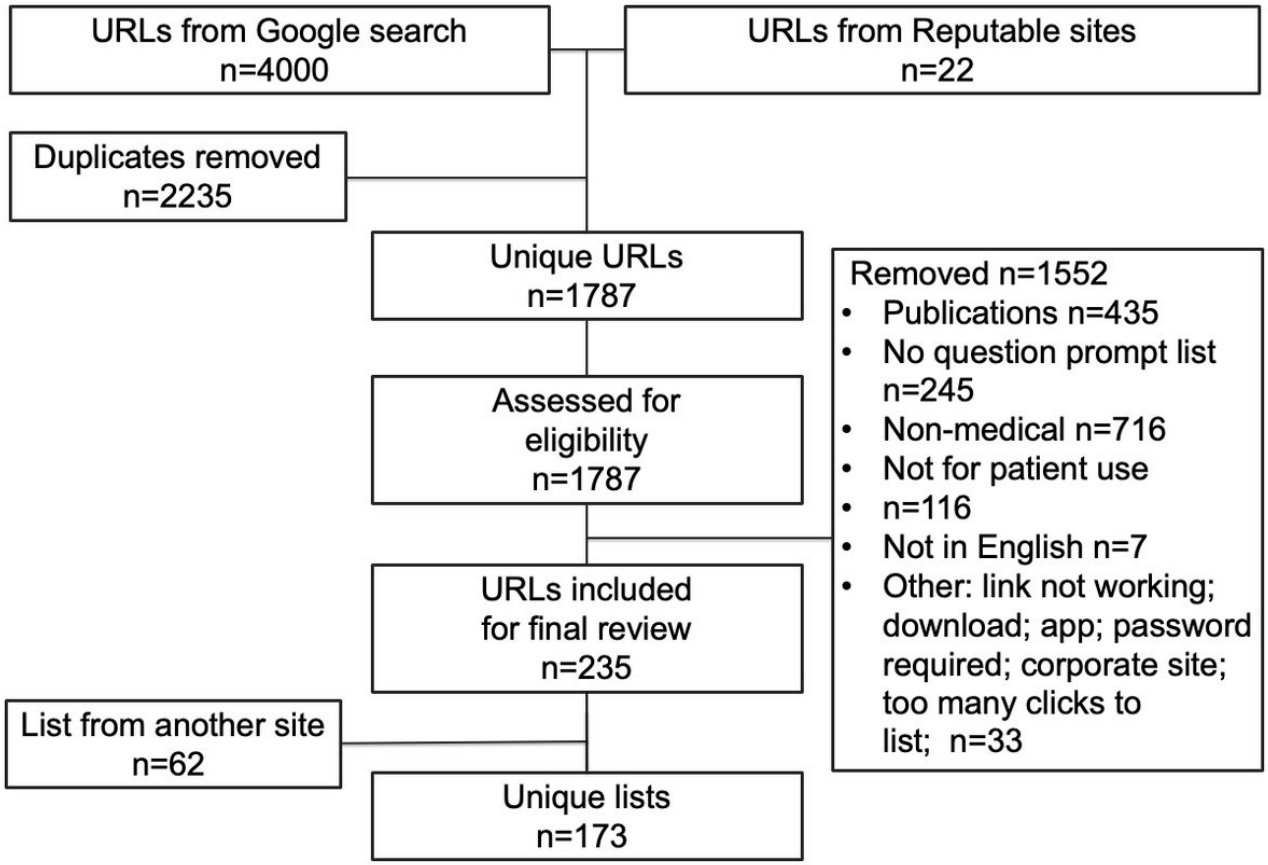


Table 4. Sample list data: readability of instructions and question lists, number of questions, source/author information, and date of creation or review.

Organization or website	Instructions and/or introduction for list		Question list		Number of questions	Evidence of source or author	Date of creation or review
	Flesch readability	FKGL[a]	Flesch readability	FKGL			
Reputable sites							
Wiser Healthcare	N/A[b]	n/a	85.5	3	9	Yes[c]	No
Cancer Australia	72.4	7.7	77.2	4.7	51	Yes[c]	Yes[d]
NPS[e] Choosing wisely	76.5	7.5	70.6	6.0	10	Yes	Yes[d]
Healthdirect	48.0	11.2	80.7	4.2	56	Yes	Yes[d]
NHS[f]	71.8	7.7	92.3	2.1	20	No	Yes[g,h]
Ask 3 Questions NHS	85.5	4.8	98.1	2.1	3	Yes[c]	No
IHI[i] Ask Me 3	64.6	9.4	100.0	0.5	3	Yes	No
Ask Share Know	78.8	5.5	75.4	5.2	3	No	No[j]
Ask 3 Questions Cardiff, UK	78.2	4.8	71.1	5.4	3	Yes[c]	No
Cancer Council (Australia)	58.7	8.7	77.4	4.6	34	Yes	Yes[h]
Average of all reputable sites	70.5	7.7	82.8	4.2	19	80% (Yes)	50% (Yes)
Other sites							
Australian Thyroid Foundation	n/a	n/a	70.1	6.3	8	No	No[i]
American Heart Association (Heart Failure)	56	10.3	82.2	3.8	31	No	No[i]
HSS[k] Orthopaedic Hospital	54.7	11.6	64.1	7.5	50	Yes	Yes[d]
Association for Children's Mental Health	59.3	9.2	78.3	5.5	30	No	No[j]
Beyond Blue	50.4	10.8	85.6	3.6	5	No	No[j]
HealthyWomen.org	73.0	7.7	71.7	5.8	18	Yes	Yes[d,h]
Readers Digest	56.6	10.5	84.9	3.5	12	Yes	Yes[l]
Psych Central	53.7	10.5	63	7.5	18	Yes[c]	Yes[d,h]
The Foundation for Peripheral Neuropathy	58.2	9.8	64	6.8	23	No	No[j]
MS[m] Trust UK	67.2	7.8	85.6	3.5	27	No	Yes[d]
Average of other websites	56.0	10.3	75.0	5.4	22	40% (Yes)	50% (Yes)
Overall	63.25	9.0	78.9	4.6	20.5	60% (Yes)	50% (Yes)

[a]FKGL: Flesch-Kincaid Grade Level.

[b]There were no instructions preceding the QPL.

[c]reference.

[d]creation.

[e]NPS: National Prescribing Service.

[f]NHS: National Health Service.

[g]original.

[h]update.

[i]IHI: Institute for Healthcare Improvement.

[j]site copyright date.

[k]HSS: Hospital for Special Surgery.

[l]online publication date.

[m]MS: multiple sclerosis.

Discussion

Principal Findings

This environmental scan of the internet for patient QPLs designed for use by patients to ask questions when seeing a doctor identified 173 unique lists on 235 websites. Of the lists found in our search, 15 lists had been used by other websites, with the majority (9/15) of these lists from our original list of reputable websites. The remaining duplicated lists were also from websites we regarded as reputable. In addition, there remained an abundance of advice to users about questions they could, or should, be asking at medical appointments from a wide variety of sources.

We noted that question lists appeared in a wide variety of types of websites, both from the categories of our reputable websites, such as government health information websites (eg, Healthdirect Australia) and disease-specific websites (eg, Cancer Council Australia and America), and from news websites (eg, globalnews.ca), commercial sites and blogs (eg, yourgpsdoc.com), sites of charities (eg, thebraintumourcharity.org), and educational institution websites (eg, sydney.edu.au). URL domain extensions do not always reflect the type of institution publishing the site and materials, yet they influence the user perception of site information [23]. While the source of medical information has not been shown to correlate with the accuracy of medical information [23,34], we noted the majority of lists came from .org extensions, which are generally associated with non-commercial organizations.

While there are detailed and extensive guidelines for the development of decision aids [17], there are currently no accepted standards that apply to QPLs. Many indicators of the quality of patient health information websites have been used in the past, with few variables showing any correlation to information quality, as measured by the accuracy or currency of information presented [34]. Display of creation and update dates, the age of the site, authorship, and URL extension have not been found useful in assessing sites. Our findings of these "quality" indicators were similar to other investigations of online patient materials in these respects. The sample QPL websites had higher rates of display of creation or update dates than those in previous studies of health information on the internet [23,34]. Display of some form of authorship or attribution of the sample materials was also higher than in similar studies of online health resources [23,34]. This is likely due to at least half of the sample being what we regarded as reputable.

In our environmental scan for online QPLs for use when attending a medical appointment, it was unclear in most cases whether doctors endorsed the questions for use in consultations. Prior research has shown that doctor endorsement significantly improves the use of QPLs by patients [4]. There were a couple of exceptions where the lists were placed on the website of particular practitioners where it could be assumed that the doctors in that practice supported patients using the questions included in the lists. In general, however, we were not able to ascertain specific endorsement by medical practitioners for the lists in practice. Further qualitative research to determine doctors' views on implementing such a wide variety of QPL resources into practice is needed.

Our results show that there is room for improvement in the readability of instructions for using online QPLs. International Patient Decision Aid Standards recommendations for reading grade level for decision aids is a less than an 8th grade level [35]. While all the QPLs in the sample we tested had a grade level that met the International Patient Decision Aid Standards recommendations, accompanying instructional text was often well above this level, especially in our "other" website category.

Strengths and Limitations

By utilizing the strengths of the environmental scan method, we were able to rapidly assess the online environment, where most Australians look for health information, for the presence of QPLs. While literature reviews have been able to show the benefits of QPLs in practice, no assessment has been made of the number and availability of similar tools in the real-world environment. The online environment changes rapidly, and this scan provides a survey at a point in time of what is available to users. Our search tool was Google.com.au; hence, a similar search strategy in other countries might yield different results. We also limited our search to sites in English.

We noted website updates occurring even as we reviewed the websites; if the scan was to be replicated in the future, the results may be different. Further assessment of the quality of the lists beyond readability would provide additional information. Some proxy measures were used to assess sites, even though the presence of, for example, a date and authorship has been shown to have little correlation with quality. It is not clear which, if any, QPLs we discovered have been evaluated for their efficacy in improving patient and doctor outcomes in consultations.

It is unlikely that Google users might search specifically for question lists using the search terms we used. We were not testing the utility of terms to find QPLs; rather, the intent of our search and choice of search terms were to find as many as possible of the question lists available to users. Research into how users actually access, assess, and use these resources is warranted.

Conclusions

People seeking information on Google.com.au have a vast number of question lists available to them to use in consultations with their doctor. Surrogate markers of the quality or usefulness of various types of online QPLs have not yet been established, which makes it difficult to assess the value of the abundance of lists. Quality measures for QPLs should follow further assessment of usage and endorsement.

Physician endorsement has been shown to be an important factor in the usefulness of QPLs, but there is little information to suggest whether the QPLs found in this scan would be endorsed by physicians. Whether these diverse QPLs are endorsed by medical practitioners warrants further investigation.

Ensuring that online QPLs and the instructions for their intended use are accessible to patients is important. To improve usage of online QPLs, instructions for their use should have an equivalent readability or at least be accessible to the majority of people in the target groups for which the lists have been created.

Acknowledgments

This study was funded by a National Health and Medical Research Council, Grant/Award Number: Centre for Research Excellence; Postgraduate Research Scholarship in the Evaluation of the Healthdirect Question Builder (SC2396).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Source and URL for QPLs identified.
[DOCX File , 31 KB - jmir_v22i5e17002_app1.docx]

References

1. Shay LA, Lafata JE. Where Is the Evidence? A Systematic Review of Shared Decision Making and Patient Outcomes. Med Decis Making 2014 Oct 28;35(1):114-131. [doi: 10.1177/0272989x14551638]
2. Härter M, Moumjid N, Cornuz J, Elwyn G, van der Weijden T. Shared decision making in 2017: International accomplishments in policy, research and implementation. Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen 2017 Jun;123-124:1-5. [doi: 10.1016/j.zefq.2017.05.024]
3. Australian Commission on Safety and Quality in Health Care. Australian Commission on Safety and Quality in Health Care. Sydney: ACSQHC; 2017. National Safety and Quality Health Service (NSQHS) Standards (second edition) URL: https://www.safetyandquality.gov.au/ [accessed 2019-11-14]
4. Sansoni JE, Grootemaat P, Duncan C. Question Prompt Lists in health consultations: A review. Patient Educ Couns 2015 Jun 03;98(12):1454-1464. [doi: 10.1016/j.pec.2015.05.015] [Medline: 26104993]
5. Shepherd HL, Barratt A, Jones A, Bateson D, Carey K, Trevena LJ, et al. Can consumers learn to ask three questions to improve shared decision making? A feasibility study of the ASK (AskShareKnow) Patient-Clinician Communication Model(®) intervention in a primary health-care setting. Health Expect 2016 Dec;19(5):1160-1168 [FREE Full text] [doi: 10.1111/hex.12409] [Medline: 26364752]
6. Tan SS, Goonawardene N. Internet Health Information Seeking and the Patient-Physician Relationship: A Systematic Review. J Med Internet Res 2017 Jan 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
7. Lee K, Hoti K, Hughes JD, Emmerton L. Dr Google and the consumer: a qualitative study exploring the navigational needs and online health information-seeking behaviors of consumers with chronic health conditions. J Med Internet Res 2014 Dec;16(12):e262 [FREE Full text] [doi: 10.2196/jmir.3706] [Medline: 25470306]
8. Australian Institute of Health and Welfare. Australian Government, 20 June. 2018. Australia's health 2018: in brief URL: https://www.aihw.gov.au/reports/australias-health/australias-health-2018/contents/table-of-contents [accessed 2020-04-28]
9. Tracy M, Shepherd H, Ivers R, Mann M, Chiappini L, Trevena L. What patients want to ask their doctors: Data analysis from Question Builder, an online question prompt list tool. Patient Educ Couns 2020 May;103(5):937-943 [FREE Full text] [doi: 10.1016/j.pec.2019.11.023] [Medline: 31831304]
10. Tan SS, Goonawardene N. Internet Health Information Seeking and the Patient-Physician Relationship: A Systematic Review. J Med Internet Res 2017 Jan 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
11. Bowel Cancer Austalia. 2017 Oct 17. What I need to ask (updated 2014) URL: https://www.bigbowel.com.au/what-i-need-to-ask [accessed 2017-10-17]
12. Cancer Council Australia. 2016 Mar 09. Questions to ask your doctor (updated March 9 2016) URL: http://www.cancer.org.au/about-cancer/after-a-diagnosis/questions-to-ask-your-doctor.html [accessed 2017-10-17]
13. St Vincent's Health Australia. 2014. Questions to ask your doctor and health care team URL: https://svha.org.au/home/our-care/being-involved-in-your-care/questions-to-ask-your-doctor-and-health-care-team [accessed 2017-10-17]
14. Walsh K, Pryor T, Reynolds K, Walker J, Mobilizing Minds Research Group. Searching for answers: How well do depression websites answer the public's questions about treatment choices? Patient Educ Couns 2019 Jan;102(1):99-105 [FREE Full text] [doi: 10.1016/j.pec.2018.08.019] [Medline: 30150127]
15. Muscat DM, Shepherd HL, Morony S, Smith SK, Dhillon HM, Trevena L, et al. Can adults with low literacy understand shared decision making questions? A qualitative investigation. Patient Educ Couns 2016 Nov;99(11):1796-1802. [doi: 10.1016/j.pec.2016.05.008] [Medline: 27344226]

16. Sørensen K, Van DBS, Fullam J, Doyle G, Pelikan J, Slonska Z, et al. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012 Jan 25;12:80 [FREE Full text] [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]
17. Elwyn G, O'Connor A, Stacey D, Volk R, Edwards A, Coulter A, et al. Developing a quality criteria framework for patient decision aids: online international Delphi consensus process. BMJ 2006 Aug 26;333(7565):417 [FREE Full text] [doi: 10.1136/bmj.38926.629329.AE] [Medline: 16908462]
18. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999 Feb;53(2):105-111 [FREE Full text] [Medline: 10396471]
19. Khazaal Y, Chatton A, Cochand S, Coquard O, Fernandez S, Khan R, et al. Brief DISCERN, six questions for the evaluation of evidence-based content of health-related websites. Patient Educ Couns 2009 Oct;77(1):33-37. [doi: 10.1016/j.pec.2009.02.016] [Medline: 19372023]
20. Reavley N, Jorm A. The quality of mental disorder information websites: a review. Patient Educ Couns 2011 Nov;85(2):e16-e25 [FREE Full text] [doi: 10.1016/j.pec.2010.10.015] [Medline: 21087837]
21. Harris PR, Sillence E, Briggs P. The Effect of Credibility-Related Design Cues on Responses to a Web-Based Message About the Breast Cancer Risks From Alcohol: Randomized Controlled Trial. J Med Internet Res 2009 Aug 25;11(3):e37. [doi: 10.2196/jmir.1097]
22. Sillence E, Briggs P, Harris PR, Fishwick L. How do patients evaluate and make use of online health information? Soc Sci Med 2007 May;64(9):1853-1862. [doi: 10.1016/j.socscimed.2007.01.012] [Medline: 17328998]
23. Bernstam EV, Walji MF, Sagaram S, Sagaram D, Johnson CW, Meric-Bernstam F. Commonly cited website quality criteria are not effective at identifying inaccurate online information about breast cancer. Cancer 2008 Mar 15;112(6):1206-1213 [FREE Full text] [doi: 10.1002/cncr.23308] [Medline: 18266210]
24. Ieong S, Mishra N, Sadikov E, Zhang L. Domain bias in web search. 2012 Presented at: Epub International Conference on Web Search and Data Mining (WSDM); February 2012; Seattle, WA. [doi: 10.1145/2124295.2124345]
25. Aggarwal S, Oostendorp H, Reddy Y, Indurkhya B. Providing Web Credibility Assessment Support. 2637260: ACM; 2014 Presented at: 2014 European Conference on Cognitive Ergonomics; 2014; Vienna, Austria p. 1-8. [doi: 10.1145/2637248.2637260]
26. Bonner C, Fajardo MA, Hui S, Stubbs R, Trevena L. Clinical Validity, Understandability, and Actionability of Online Cardiovascular Disease Risk Calculators: Systematic Review. J Med Internet Res 2018 Feb 01;20(2):e29. [doi: 10.2196/jmir.8538]
27. Fajardo MA, Balthazaar G, Zalums A, Trevena L, Bonner C. Favourable understandability, but poor actionability: An evaluation of online type 2 diabetes risk calculators. Patient Educ Couns 2019 Mar;102(3):467-473. [doi: 10.1016/j.pec.2018.10.014] [Medline: 30389187]
28. statcounter. 2018. Search Engine Market Share in Australia - December 2018 URL: http://gs.statcounter.com/search-engine-market-share/all/australia/2016 [accessed 2019-01-08]
29. Kinnersley P, Edwards A, Hood K, Cadbury N, Ryan R, Prout H, et al. Interventions before consultations for helping patients address their information needs. Cochrane Database Syst Rev 2007 Jul 18(3):CD004565-CD004568 [FREE Full text] [doi: 10.1002/14651858.CD004565.pub2] [Medline: 17636767]
30. Kim J, Thomas P, Sankaranarayana R, Gedeon T, Yoon H. Eye-tracking analysis of user behavior and performance in web search on large and small screens. J Assn Inf Sci Tec 2014 Jun 11;66(3):526-544. [doi: 10.1002/asi.23187]
31. Singer G, Norbisrath U, Lewandowski D. Ordinary search engine users carrying out complex search tasks. Journal of Information Science 2012 Dec 17;39(3):346-358. [doi: 10.1177/0165551512466974]
32. Flesch R. A new readability yardstick. J Appl Psychol 1948 Jun;32(3):221-233. [doi: 10.1037/h0057532] [Medline: 18867058]
33. Kincaid JP, Fishburne RP, Rogers RL, Chissom BS. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. Institute for Simulation and Training 1975;56:1-49.
34. Buhi ER, Daley EM, Oberne A, Smith SA, Schneider T, Fuhrmann HJ. Quality and accuracy of sexual health information web sites visited by young people. J Adolesc Health 2010 Aug;47(2):206-208. [doi: 10.1016/j.jadohealth.2010.01.002] [Medline: 20638014]
35. International Patient Decision Aid Standards (IPDAS) Collaboration. IPDAS 2005: Checklist for judging the quality of patient decision aids URL: http://www.ipdas.ohri.ca/using.html2006 [accessed 2017-10-17]

Abbreviations

ADHD: attention deficit hyperactivity disorder.
FKGL: Flesch-Kincaid Grade Level.
FRE: Flesch Reading Ease.
HSS: Hospital for Special Surgery.
IHI: Institute for Healthcare Improvement.

MS: multiple sclerosis.
NHS: National Health Service.
NPS: National Prescribing Service.
NSW: New South Wales.
QPL: question prompt list.
SDM: shared decision making.

Edited by G Eysenbach; submitted 11.11.19; peer-reviewed by Z Michaleff, T Wieringa; comments to author 16.12.19; revised version received 13.02.20; accepted 23.03.20; published 29.05.20.

Please cite as:
Tracy MC, Shepherd HL, Patel P, Trevena LJ
Exploring the Vast Choice of Question Prompt Lists Available to Health Consumers via Google: Environmental Scan
J Med Internet Res 2020;22(5):e17002
URL: http://www.jmir.org/2020/5/e17002/
doi:10.2196/17002
PMID:32469321

Original Paper

Adopting Patient Portals in Hospitals: Qualitative Study

Pauline Hulter[1], MSc; Bettine Pluut[1], PhD; Christine Leenen-Brinkhuis[2], MSc; Marleen de Mul[1], PhD; Kees Ahaus[1], PhD; Anne Marie Weggelaar-Jansen[1,3], MCM, PhD

[1]Erasmus School of Health Policy & Management, Erasmus University Rotterdam, Rotterdam, Netherlands

[2]Dutch Hospital Association, Utrecht, Netherlands

[3]Eindhoven University of Technology, Eindhoven, Netherlands

Corresponding Author:
Pauline Hulter, MSc
Erasmus School of Health Policy & Management
Erasmus University Rotterdam
Burgemeester Oudlaan 50
Rotterdam, 3062 PA
Netherlands
Phone: 31 618406943
Email: hulter@eshpm.eur.nl

Abstract

Background: Theoretical models help to explain or predict the adoption of electronic health (eHealth) technology and illustrate the complexity of the adoption process. These models provide insights into general factors that influence the use of eHealth technology. However, they do not give hospitals much actionable knowledge on how to facilitate the adoption process.

Objective: Our study aims to provide insights into patient portal adoption processes among patients and hospital staff, including health care professionals (HCPs), managers, and administrative clerks. Studying the experiences and views of stakeholders answers the following question: How can hospitals encourage patients and HCPs to adopt a patient portal?

Methods: We conducted 22 semistructured individual and group interviews (n=69) in 12 hospitals and four focus groups with members of national and seminational organizations and patient portal suppliers (n=53).

Results: The effort hospitals put into adopting patient portals can be split into three themes. First, inform patients and HCPs about the portal. This communication strategy has four objectives: users should (1) know about the portal, (2) know how the portal works, (3) know that action on the portal is required, and (4) know where to find help with the portal. Second, embed the patient portal in the daily routine of HCPs and management. This involves three forms of support: (1) hospital policy, (2) management by monitoring the numbers, and (3) a structured implementation strategy that includes all staff of one department. Third, try to adjust the portal to meet patients' needs to optimize user-friendliness in two ways: (1) use patients' feedback and (2) focus on optimizing for patients with special needs (eg, low literacy and low digital skills).

Conclusions: Asking stakeholders what they have learned from their efforts to stimulate patient portal use in hospitals elicited rich insights into the adoption process. These insights are missing in the theoretical models. Therefore, our findings help to translate the relatively abstract factors one finds in theoretical models to the everyday pragmatics of eHealth projects in hospitals.

(J Med Internet Res 2020;22(5):e16921) doi:10.2196/16921

KEYWORDS

patient portal; adoption; adoption processes; eHealth

Introduction

Electronic health (eHealth) technology is generally considered promising for improving both the well-being and health of patients and the efficiency and effectiveness of the health care organization [1]. However, several studies show that its promise is not always fulfilled [1] and the results on the benefits gained are diverse [1,2]. This also applies to patient portals [2], which have sparked the interest of researchers, government policy makers, and health care organizations.

Previous studies define a patient portal in various ways [3-7]. Some consider them the same as a personal health record (PHR) [3]. Others regard them as a class of PHR [4]: whereas health care organizations own and control patient portals, PHRs are owned and controlled by patients themselves [5]. Grünloh et al

[6] define patient portals as "provider-tethered applications that allow patients to access, but not to control, certain health care information (eg, their EHR [electronic health record]) and provide communication and administrative functions (eg, secure messaging, appointment booking, and prescription refill requests)." Wildenbos [7] adds the possibility of authorizing informal caregivers to share access to patient portals.

Despite its technological focus and aim, the success or failure of a patient portal relies on how it is used by patients and staff, such as health care professionals (HCPs), managers, and administrative clerks [1,2,8]. A systematic review shows often-limited use by patients and HCPs for seven nontechnical reasons [2]: (1) patients worry about the confidentiality of their personal health data, (2) patients are unaware of the portal, have no digital access, or think it will not be useful, (3) patients have low health literacy or find using the portal too complicated, (4) HCPs worry about increased workload and disruptions to their usual tasks, (5) HCPs lack the digital skills to interact with patients, (6) HCPs worry that they cannot respond fast enough to patients' questions, and (7) HCPs are concerned that they can be held liable [2]. All seven reasons hinder the adoption of patient portals [2].

Researchers have developed theoretical models to explain or predict the adoption of eHealth technologies, including patient portals. Two systematic reviews on information technology adoption both mention three frequently used acceptance and adoption models [9,10]. Strikingly, these three general models are applied in all societal domains, not just health care [9]. Table 1 details these most-used models to facilitate an understanding of their ideas on the adoption process [9-20]. The table also includes the recently developed NASSS (nonadoption, abandonment, scale-up, spread, and sustainability) framework [21,22].

Table 1. Theoretical acceptance and adoption models.

Model	Developer	Year	Theoretical basis	Central constructs	Points of critique	Extended models
TAM[a]	Davis [9,11,12]	1985	Cognitive psychology [13]	Describes elements to predict the degree to which a person plans to perform specific future behavior. It suggests that perceived usefulness, perceived ease of use, and attitude (ie, intention to use) can explain user motivation [12]. It is a way to predict the intended use of a technology.	Mainly conceptualized for the acceptance of individuals and is not useful for explaining acceptance of electronic health (eHealth) technologies by organizations [11,14].	TAM2 (by Venkatesh and Davis) and TAM3 (by Venkatesh and Bala) [9,12]
DOI[b]	Rogers [9]	1995	Diffusion research [18]	Explains the characteristics of innovation. Observability, trialability, complexity, relative advantage, and compatibility are the primary determinants of innovation diffusion, which help explain the different rates of adoption [9,15]. Diffusion starts with recognizing the user's need. It spreads by knowledge acquisition, persuasion, decision (ie, adopt or reject), implementation (ie, routine use, reinvention, and conformation), promotion, and evaluation [15].	There is little focus on the organizational context [14,16,17].	The unifying theoretical model of Greenhalgh et al [16] and the Consolidated Framework for Implementation Research (CFIR) of Damschroder et al [17,18]
UTAUT[c]	Venkatesh et al [9,19]	2003	Cognitive psychology [13]	Builds on TAM and focuses on perceptions and assumptions of people, resulting in the intention to use technology. States that constructs like performance expectancy, effort expectancy, social influence, and facilitating conditions influence intention and ultimately behavior [19]. These four constructs are moderated by gender, age, experience, and voluntariness of use [19].	Does not deal with hindrances to actual use [13]. Excludes users' cognitive, affective, and physical ability to use technology [20] and ignores technological factors that might influence the decision to use an application [20].	UTAUT2 by Venkatesh et al developed in 2012 [9]
NASSS[d] framework	Greenhalgh et al [21,22]	2017	Complexity theory [21,22]	Points to aspects explaining the complexity of technological innovations in health care, which according to all the described models influence the adoption. Includes the value proposition (ie, supply-side and demand-side values) as an important factor, in contrast to many implementation theories that do not [14,21].	Not found yet	Not found yet

[a]TAM: Technology Acceptance Model.

[b]DOI: diffusion of innovations.

[c]UTAUT: Unified Theory of Acceptance and Use of Technology.

[d]NASSS: nonadoption, abandonment, scale-up, spread, and sustainability.

All four theoretical models include two key concepts—acceptance and adoption—which are either ill-defined or used interchangeably. The concept of acceptance focuses on if, how, and when intended users would use a technology [23], and adoption is the actual use of an eHealth technology. Different stages in an adoption process can result in the actual use of an eHealth technology [16]. During an adoption process, users of eHealth technology develop feelings about the technology, gain experience, find meaning or do not find meaning in its use, and evaluate the functions of the technology [16]. Thus, there is a difference between intended use (ie, acceptance) and actual use (ie, adoption) [13].

The complexity of the adoption process of eHealth technologies is underexposed in all four theoretical models. The literature reports that adoption is a highly complex process and that results cannot be *made* or even predicted [13,16]. Greenhalgh et al highlight specific prerequisites for each of the three adoption process stages for innovations [16]: (1) in the preadoption stage, intended adopters should know about the innovation: in this case, the patient portal, (2) in the early use stage, intended adopters should be supported in using the innovation and learn how to fit or blend it into their daily routines, (3) in the adoption stage, established users arise if they gain an understanding of the consequences of using of the innovation and if they have the opportunity to refine and improve it: in this case, a patient portal [16]. The theoretical models provide no clarity on the pragmatics of efforts in the adoption process. For instance, they do not show how hospitals can encourage patients and HCPs to adopt a patient portal. Therefore, we studied the introduction of patient portals in 12 Dutch hospitals, using a multi-actor perspective to gain a broad understanding of the experiences and views on adoption. Our empirical study focuses on the pragmatics of stimulating the adoption of a patient portal. This paper answers the following research question: How can hospitals encourage patients and HCPs to adopt a patient portal?

Methods

Design and Setting

In this qualitative study, we asked participants from various backgrounds how they encouraged users to adopt a patient portal and what they had learned from their efforts, in order to understand what stimulated or hindered the adoption process. We conducted 22 multi-actor, semistructured group interviews and held four structured focus groups to check, enlarge, and enrich our findings [24-26]. Table 2 lists the different forms of data collection.

Table 2. Data collection.

Type of data collection and participants	Number of participants (N=122), n (%)	Data collection moments (N=26), n (%)
Individual and group interview		
Patients	22 (18.0)	22 (85)
Health care professionals	16 (13.1)	22 (85)
Organizational staff	31 (25.4)	22 (85)
Focus group		
Project leaders and staff 1	14 (11.5)	2 (8)
Project leaders and staff 2	28 (23.0)	2 (8)
Patient portal suppliers	5 (4.1)	1 (4)
Macro stakeholders	6 (4.9)	1 (4)

All interviews, both group and individual, and focus group sessions took place in a hospital or online. All the hospitals included in this study are participating in a national program—VIPP (Versnellingsprogramma Informatie-Uitwisseling Patiënt en Professional) [27]—initiated by the Dutch Hospital Association and the Dutch government. VIPP is the Dutch government's financial incentive program to support information exchange between patients and professionals through patient portals. The aim of the VIPP program is to give patients online access to their medical data, either through a patient portal or a PHR [27]. Information technology (IT) suppliers with a commercial interest deliver patient portals and PHRs. Given that Dutch hospitals are free to choose any supplier for their patient portals or PHRs, the hospitals in this study use different patient portals. The portals might differ in their available functionalities, but they all offer patients online access to their personal health information [6]. Additionally, hospitals had different aims for their patient portals; only the VIPP program aims were similar for all Dutch hospitals. Hospitals receive financial support based on their achievement of specific national VIPP aims; for example, "In the past 30 days, 25% of all patients (based on DRG [diagnosis-related group] contacts) logged in to the patient portal or the link to a PHR" [28]. How hospitals achieve these aims is left up to the hospital.

Recruitment and Participants

There are three categories of Dutch hospital: academic, teaching, and general. Academic hospitals were excluded from our study because at the time they were not participating in the VIPP program. We first determined inclusion criteria for general and teaching hospitals, aiming for a diverse study group. Based on user statistics of patient portals (low and high), geographical differences (rural areas and cities), variation in patient portal suppliers, and usage of the patient portal (more or less than one year), the researchers (PH, AMWJ, and BP) made a selection of targeted hospitals.

In total, 15 of 64 Dutch hospitals (23%) were approached, of which 12 (80%: 10 teaching and 2 general) agreed to participate. Reasons for not participating included "already participating in another study" (n=1) and "too busy with implementation and fear of not meeting VIPP deadlines," which would mean losing financial support (n=2).

At each hospital or via the Zoom online platform [29], the researchers arranged individual and group interviews in close collaboration with the person running the implementation of the patient portal at that particular hospital. Most often, this person was the project leader who selected participants according to a predefined list of three roles:

1. Patients: patients and client council members (n=22).
2. HCPs: physicians, Chief Medical Information Officers (CMIOs), nurses, Chief Nursing Information Officers (CNIOs), pharmacists, and outpatient clinic staff (n=16).
3. Organizational staff: project leaders, project staff, communication advisors, legal policy makers, and managers (n=31).

In total, 69 participants were included (see Table 2) if they were older than 18 years and had experience with developing and/or using a patient portal. The researchers emailed invitations to participants of the individual and group interviews; groups ranged from 2 to 6 participants. In 10 of the participating hospitals, we organized one or two group interviews; in the remaining two hospitals we conducted one individual interview for logistical reasons.

The aim of the focus groups was to check, enlarge, and enrich our results. Two of the four focus groups were held with project leaders and project staff of hospitals. They joined one of two self-selected focus groups organized during an educational meeting of the VIPP program (n=14 and n=28, respectively). For the other two, we used targeted sampling, selecting experts from patient portal suppliers (n=5) for the third group. The fourth group included macro stakeholders: Ministry of Health, NICTIZ (Nationaal ICT Instituut in de Zorg), the center of expertise for eHealth, health insurance companies, and scientific experts (n=6). All focus group members were invited to take part by email.

Data Collection and Analysis

Qualitative data were collected on-site in the hospital (n=59) or online via the Zoom platform [29] (n=10) in the fall of 2018. The Zoom platform enabled the inclusion of hard-to-reach, geographically dispersed participants [30]. Group interviews lasted an average of 72 minutes (range 53-88) and the individual interviews lasted an average of 53 minutes (range 44-65). One researcher (PH) conducted all the interviews, following a predefined topic list (see Multimedia Appendix 1) [31] that was based on a search of the literature and discussed among the research team.

The four focus groups lasted an average of 82 minutes (range 71-88) and were steered by a Microsoft PowerPoint presentation explaining the findings of our study. No revisions were made in the presentation between focus group sessions, ensuring that varying opinions were heard before conclusions were drawn [32]. Each focus group was run by two researchers (PH and BP or PH and AMWJ) complementing each other: one moderating and the other taking notes.

During both group interviews and focus group sessions we encouraged the exchange of heterogeneous views that provided insights into similarities and differences in the opinions and experiences of the various stakeholders [26,30,31]. We also invited the participants to challenge each other's views [33], to explore the implications of their thinking, and to articulate their sometimes-implicit assumptions about the adoption process. This method generated new insights through group interactions.

We audio-recorded the on-site interviews and focus groups and video-recorded the online sessions. All interviews were transcribed verbatim. Analysis, comprising six phases [34], was not linear but a recursive process. First, each individual researcher gave the transcripts a close reading. Second, one researcher (PH) developed codes for the interesting parts of the data. Next, three researchers (PH, AMWJ, and BP) independently developed themes, reaching consensus on a list of relevant themes (eg, communication channel, ambassadors, and patient participation [35]) through discussion. Fourth, one researcher (PH) read the transcripts again. Fifth, using the list of themes, one researcher (PH) performed thematic coding, which two other researchers (AMWJ and BP) checked. In the final phase we (PH, AMWJ, and BP) started writing the results [34].

Ethical Approval

The Medical Ethics Review Committee of Erasmus Medical Center (Erasmus MC) approved our research proposal (MET-2018-1531) and checked if we were General Data Protection Regulation (GDPR) compliant. All participants were asked to sign an informed consent form.

Results

Overview

Our study aims to provide insights into the adoption process of a patient portal by patients and HCPs. The efforts of hospitals to stimulate the use of patient portals can be categorized under three themes: (1) informing patients and professionals about the patient portal, (2) embedding the patient portal in the daily routine of HCPs, and (3) adjusting the portal to patients' needs in the initial and continuous development process. Below we present our empirical findings for each of these themes.

Informing Patients and Health Care Professionals About the Patient Portal

Overview

Participants agreed that communication about the patient portal is very important for the adoption of patient portals. The hospitals require the involvement of communication experts to inform patients and HCPs. We identified four objectives to support patient portal adoption using the informing of patients and HCPs as the basis. Here we explain them one by one. Participants mentioned using 23 communication channels to reach their audiences (see Table 3).

Table 3. Channels (N=23) used to achieve four communication objectives to inform patients and health care professionals.

Channel	Knowing about the portal	Knowing how the portal works	Knowing that action is required on the portal	Knowing where to find help with the portal
Pocket tickets	x[a]			
Promotion leaflets	x			
Press releases	x	x		
Posters	x			
Banners	x			
Screen in waiting room	x	x		
Social media	x			
Video on website	x	x		
Explanatory leaflets	x	x		
Students or volunteers in central hall for a week	x	x		x
Physical point with employees and volunteers	x	x		x
Health care professionals	x	x	x	x
Letter with request (before or after appointment)			x	
Text message with request			x	
Mail with request			x	
Staffed desk				xxx
Informative meetings	xx			
Interactive meetings	xxx	xxx		
Internal website	xx			
Peer to peer	xx	xx	xx	xx
Training	xx	xx	xx	
Newsletter	xx			
Goodie bag with explanation	xx	xx		

[a]Key: x = patients; xx = professionals; and xxx = patients and professionals.

Objective 1: Knowing About the Portal

First, participants argued that it is obviously necessary to make sure that patients and HCPs know about the patient portal; otherwise, they cannot use it. Meeting this objective means that patients and HCPs will have a general idea of what the portal is and what it looks like. The hospitals used various mass-focused channels most frequently to communicate the existence of the portal to patients and HCPs. These channels include leaflets and posters, banners hung in the corridors, messages on social media, and placing volunteers in the central hall of the hospital to point patients to the portal. Using multiple channels to reach patients is considered important because patients have different preferences, but it is probably a bit inefficient, as these quotes illustrate:

> *I am not inclined to take leaflets from the hospital. If it's really important, I think I will be reminded.* [Participant #0.4, client council member]

> *All those freebies [goodie bags], I don't like them. They won't get me to look at the portal.* [Participant #14.2, patient, informal caregiver, and former client council member]

> *Still, it works for others.* [Participant #11.4, client council member]

Looking at the channels used to reach HCPs, we see hospitals organizing informative presentations for medical and nursing staff and department meetings. Participants felt that informing only the team leaders and managers is not enough; it is important to directly inform the HCPs. According to participants, effective channels that reach broad groups of professionals include department meetings, the hospital intranet, and the staff newsletter. Reaching out by email is considered inefficient because HCPs receive a lot of email and particular messages can be easily missed or skipped.

Objective 2: Knowing How the Portal Works

Second, participants argued that patients and professionals need to know how the patient portal works. Hospitals can meet this objective by (1) using video screens in waiting rooms, (2) putting explanatory videos on the hospital website, and (3) placing volunteers in the central hall of the hospital to teach patients how the portal works.

Hospitals asked professionals who are already successful portal users to explain how the portal works to their peers (ie,

peer-to-peer information). One hospital gave their professionals a goodie bag with explanatory flyers during portal training, but only a few hospitals organized training sessions on new functionalities for HCPs. However, the participants said that HCPs need to know how the portal actually works as patients sometimes turn to them for help with portal questions:

> *So often I hear that people can't log in to the patient portal. I think that as staff we should be looking into the portal far more. How does it actually work for the patient in practice?* [Participant #1.2, staff member]

If the HCP has to tell the patient that they do not know how to help, this can be a disappointment. One participant reflected the following:

> *Lots of patients say, "I'm logged in but then I don't know [what to do]." Many colleagues say, "I don't know either." If you can't help the patient, they'll drop out immediately.* [Participant #5.5, senior doctor's assistant]

Objective 3: Knowing That Action Is Required on the Portal

A third objective is to entice patients and HCPs to visit the portal. Hospitals did this by sending specific calls to action, including personal letters, text messages, or emails asking the recipient to read through an online brochure on their treatment before their hospital appointment or to fill in a questionnaire:

> *Your appointment letter also asks you to complete the pre-operative screening questionnaire at home.* [Participant #9.1, communication advisor]

Hospitals encouraged HCPs to use the patient portal by asking them to respond to e-consults (ie, messages sent by the patient) and by showing them functions that will save their time or make their work more efficient. Time-saving functions like preoperative screening questionnaires on the patient portal are easy to embed in the daily routine of professionals. One participant explained the following:

> *We built the pre-operative screening questionnaire in such a way that [the information the patient provides] gets entered directly into the anesthesiologist's outpatient file. [This means] that the moment the anesthesiologist starts the consultation with the patient, the data are already in the system. The outpatient clinic started using the questionnaire right from the go-live.* [Participant #12.1, project leader]

Objective 4: Knowing Where to Find Help With the Portal

The fourth communication objective is to ensure that patients and HCPs know where to look for help when they have a problem using the portal. Hospitals organized a help desk, publishing its phone number and email address in leaflets, letters, on the hospital website, and on the portal itself. Some hospitals organized a service point clearly visible in the central hall of the hospital, where patients receive face-to-face service. Outpatient clinic staff can tell patients about this service point. HCPs can ask their colleagues for help (ie, peers).

Reviewing these hospital communication strategies, three findings are worth a mention:

1. The texts used to inform patients and HCPs have a promotional tone. Hospitals choose to stress the benefits and hardly mention the potential disadvantages or risks of portal use.
2. Mass communication is preferred because it is less labor intensive. However, it is also less personal and less in line with the needs of an individual. Personal communications, such as letters, text messages, and emails, have the advantage in that they probably make patients and health care professionals feel personally addressed.
3. Hospitals struggle with the timing of starting their communication efforts, for example, having a silent "go-live" or starting a campaign directly after launching the portal.

Our participants explained that most hospitals do not inform patients about the go-live, because this gives them the opportunity to solve start-up problems and technical issues signaled by the first users. After some two to three months they will use a cross-media approach to communicate on the portal. Our participants expressed concern that if a hospital initiated a communication campaign straight after launching the portal, the hospital could be making promises that they cannot keep.

Embedding the Patient Portal in Daily Routines

Overview

In the previous section we showed how communication strategies support patient portal adoption processes. HCPs worry that using the portal is time-consuming and will disturb their daily routine. They regard explaining how the portal works and communicating with patients on the portal as extra tasks and expect that helping non-computer-literate patients (eg, the elderly and people with low literacy) will be especially time-consuming. One participant explained the following:

> *If a patient says, "I don't use the computer," I won't ask if they have a son or daughter who'd like to share their access. I don't start with that, it costs too much time, really.* [Participant #5.5, senior doctor's assistant]

Our data show three ways to support embedding the patient portal in the daily routine of HCPs and management: (1) hospital policy, (2) management by monitoring the numbers, and (3) a structured implementation strategy that includes all employees in the department, termed a specialism-focused approach. According to our participants, all three ways require changing work processes and routines.

1: Hospital Policy

The first way to embed patient portals is by developing hospital policy on digitalization. Our participants revealed that some hospitals lack hospital-wide agreements, resulting in a lack of coherence for the patient. One such hospital-wide agreement could set the maximum time that HCPs have to respond to an

e-consult. According to our participants, the CMIO and the CNIO can play an important role in setting hospital-wide agreements and explaining new work routines to their colleagues (ie, peer influence). Our participants also said that some hospitals set no obligation or targets to use the portal:

> *How can we make sure the patient portal lands properly in the outpatient clinic? Good question... Well, it may have landed, but there is still no obligation [to use it] and that's the real problem.* [Participant #1.1, communication advisor]

For example, it is not clear within which time frame HCPs need to respond to patients' e-consults or how many patients per specialism should be using the portal. Responses to patients should be prompt, and the professionals need time to incorporate their responses into their daily work processes on the portal. One participant explained the following:

> *If you have to explain something to the patient on the portal or send them an email within four hours, then we need to think about how to get that complex planning process in the system. The professionals need time to think about it too. So that's development; and you need even more time to implement.* [Participant #13.2, CNIO]

2: Management by Monitoring the Numbers

The second way to embed patient portals is by ensuring that management monitors information on portal use or response time. Most hospitals lack management control of portal use. Some hospitals, however, monitor the numbers of new patient accounts and users on a dashboard or monitoring system. In one hospital, outpatient clinic employees must ask all patients presenting themselves at the desk if they are interested in opening a patient portal account. Hospital management uses this monitoring information to talk with staff who do not seem to be encouraging enough patients to sign up for the portal. One participant revealed the following:

> *Staff must register whether or not they have asked if a patient is interested in having a portal account. Now we can run reports on the employee level... We do that sometimes and then we can see that, say, Marie scores 100% on "not interested." Of course, ... then you'd have to start the conversation.* [Participant #5.4, care and operations manager]

Our participants said that fear is possibly a reason why HCPs do not motivate patients to sign up for the portal. Professionals need to overcome their own unfamiliarity and prejudices by experiencing the benefits of the patient portal. One participant declared the following:

> *Of course, we do it for the patient, but let's see where it helps the physician. Then you'll get them to at least use the patient portal.* [Participant #8.7, internist]

3: Specialism-Focused Approach

A third way to embed portals in daily routines is to apply a specialism-focused approach—a structured implementation strategy—that includes all the staff in the department. This involves a multidisciplinary project team (eg, communication advisor and project leader), management, and HCPs temporarily collaborating on changing work processes to benefit the incorporation of the patient portal into daily routines. Not trying to convince just one physician, but the whole department (eg, the outpatient clinic team), makes it easier to embed the patient portal. Working closely with project staff gives the HCPs support that is based on their needs or wishes. This approach requires giving HCPs the time to discuss their problems, share their experiences, and experiment. One participant explained the following:

> *Using this approach, we've really looked in depth at the points where the portal can be embedded better in their work process. For example, we've supported the specialism of rheumatology. They have very clear ideas about using the portal. Now we'll work actively with the health care professionals in the coming period to increase the use of the patient portal within their specialism.* [Participant #11.2, communication advisor]

According to our participants, using this approach supports giving professionals an understanding of how the portal works and how they can use it in their daily routines. However, they said that it is labor intensive for everyone involved, which slows down the adoption process hospital-wide.

Another implementation strategy is for hospitals to start off the portal adoption processes with keen, intrinsically motivated HCPs. Hospitals put effort into these professionals. They are seen as ambassadors, as *game changers*, who will convince other HCPs by setting a good example. One participant said the following:

> *I believe that starting out with the enthusiastic specialists is the most successful strategy and that's why we're starting with people who want it. We're not setting out with the difficult ones who don't want it.* [Participant #11.3, CMIO]

Adjusting the Portal to Meet Patients' Needs

Overview

Our participants said that in the continuous development process, adjusting the portal to meet the patients' needs is important. For example, enhancing user-friendliness ensures repeated use of a portal. As well, asking patients for feedback on the portal can reveal points of improvement that the project staff might not spot, as our participants explained:

> *An example: a patient tests the portal, first on a dummy and later on their own file. Someone remarks: "It's in chronological order, but the most recent is at the bottom. Why don't you put the most recent at the top of the page?"* [Participant #4.2, project leader]
>
> *It's as simple as that. You don't notice that when you are so involved.* [Participant #4.1, advisor of functional management]

According to our participants, another reason they find it important to adjust to patients' needs is because the perspectives of the patient and the communication advisor may differ:

> *I'm against all those abbreviations... Why not explain what they are? I guarantee you that half the patients won't know what the abbreviation means. Add an abbreviation list.* [Participant #10.6, patient and client council member]
>
> *It's my choice. I can write the term in full, I can explain it the abbreviation. But when I write it out completely, it becomes a very long sentence.* [Participant #10.2, communication advisor]

Our participants mentioned two methods hospitals use in the effort to optimize portal user-friendliness: (1) patient feedback and (2) focus on optimizations for patients with special needs (eg, low literacy, visually impaired, and low digital skills).

1: Patient Feedback

The first way to adjust to patients' needs in the continuous development of patient portals is to set up a panel of patients to act as a sounding board or to survey patients on their experiences and wishes. One participant reported the following:

> *We have a panel of 150 people. We sent these people a questionnaire on the patient portal and how they would like to use it.* [Participant #3.2, client council member]

Another way is to organize sessions with patients to test portal functionalities (eg, access to data, an e-consult, and filling in questionnaires). To illustrate, one hospital organized a test session for feedback and observations:

> *We invited a few patients from our patient panel. We gave them a test version of the portal and asked them to do a few assignments and fill in a questionnaire. For example: look at the patient portal and see if you get it. Give as much feedback as possible about the things that could be improved... There was one-on-one guidance. We had a large number of employees involved, so that we could sit next to the patients and get as much feedback as possible. So, we could also see how things went.* [Participant #7.1, project employee]

Hospitals also asked for feedback and reused questions, comments, and complaints patients express to the helpdesk. One participant said the following:

> *We now actively request feedback from patients. The helpdesk also receives feedback and phone calls and we can use the input obtained.* [Participant #1.1, communication advisor]

2: Focus on Optimizations for Patients With Special Needs

A second way to adjust to patients' needs is by optimizing the portal for people with special needs. For example, language experts check the language used on the portal and written information on how to use it, removing jargon and abbreviations and simplifying texts for patients with low literacy. They make more use of visuals (eg, icons, pictograms, and infographics):

> *You can summarize in pictograms, which makes it much easier for patients with low literacy. Visuals work better and faster.* [Participant #9.2, patient]

Another example of optimizing portal use for people with special needs is when hospitals collaborate with organizations offering general computer courses, such as the municipality, community centers, and libraries, for patients with few digital skills. Hospitals ask those organizations to blend the patient portal into their course and teach patients to work with it. Also, hospitals may refer patients to this course if they do not have computer skills and need to learn how to work with the Dutch national identity authentication method (DigiD). Participants report that the DigiD is not easy to use and its log-in process requires many steps:

> *I find the accessibility of the patient portal a real problem. Logging in with your DigiD is difficult.* [Participant #3.3, functional manager]
>
> *We look for courses on using the DigiD subsidized by the municipality. They organize courses in the community centers for people having trouble with DigiD and then these people can practice logging in on the patient portals.* [Participant #11.2, communication advisor]

Despite the importance of adjusting to patients' needs in continuous development, hospitals sometimes hesitate to include patients, because they may not be able to act on the patients' feedback. For example, if patients miss functionalities, it can require time and money to add them to the portal and, therefore, this cannot be easily fixed. Participants mentioned the importance of explaining to patients what the hospital does with their feedback and why some feedback points cannot be solved in the near future (eg, technologically impossible or too expensive). Otherwise, patients will feel that the hospital is not taking their feedback seriously. The following quote shows how hospitals struggle to let patients participate in the continuous portal development, even though they find patient input invaluable:

> *If we invite the panel group for testing, then we have to show that we have improved the portal based on their feedback... Otherwise they will think "nothing happens with our input." If we organize patient participation, you can only say "we're too busy" once.* [Participant #4.1, advisor of functional management, and participant #4.2, project leader]

One of the challenges of acting on patient feedback stems from the collaboration with the suppliers of patient portals. Suppliers will undertake to improve or develop new functionalities when multiple hospitals make the same request. Surprisingly, however, suppliers (n=5) said that they include no patients in their development process. The suppliers see it as the responsibility of their customers—the hospitals—to give voice to patients' wishes.

Discussion

Principal Findings

This qualitative study focuses on patient portal adoption processes by patients and HCPs in a Dutch hospital context. Overall, our results show that the adoption of patient portals is more dynamic than presented in theoretical models and the literature. Greenhalgh et al's linear adoption stages (ie, preadoption, early use, and established users) [16] seem useful in studying adoption by individuals, but hospital patients and HCPs are in different adoption stages. Consequently, an organization cannot simply move through sequential stages; it needs ongoing effort to be put into informing, embedding, and adjusting to patients' needs. Their focus on individuals rather than the organizational context is also a criticism levied at the theoretical models (see Table 1) [14,16,17].

All participating hospitals seem to be experimenting with stimulating adoption of the patient portal. They are trying to create effective communication strategies, looking for the best way to embed the portal in daily routines and adjust to its patients' needs. As yet, they have not found the best way of encouraging portal use by patients and HCPs. Here we explain the implications of our results.

Our study shows that hospitals are experimenting with many communication channels (N=23), mostly ones that are already in use. Despite efforts by communication departments, it seems that portal adoption is still quite a challenge. It seems that hospitals do not know which channels are most effective for which target audiences and what the right timing is for their communication campaigns. Looking at their communication strategies, we found that hospitals choose to emphasize the benefits of portal use and hardly mention the potential disadvantages or risks. According to Greenhalgh et al [16], intended adopters must know the consequences of adopting a patient portal to become established users. If intended adopters are not informed of the potential disadvantages, then they cannot oversee all the consequences of using the portal, for example, the risks. However, the financial incentives of the VIPP program may explain the positive promotion strategy. If hospitals do not attain a certain adoption percentage (ie, 10% or 25%) they will have to repay their VIPP grant.

Our results show that hospitals invest in HCP adoption through peer-to-peer influence. However, focusing on the enthusiastic HCPs can mean that the less-motivated HCPs will lag behind. That a patient portal often does not reduce the burden of HCPs (ie, it only means extra work) and that it is not embedded properly in work routines can hinder adoption. The specialism-focused approach offers a way of encouraging patient portal adoption by HCPs. This experiment with portal embedding would be interesting to study in other contexts to see where and how it could lead to better embedding of the patient portal. A possible disadvantage of this approach could be that patients will not understand why specialisms are in different adoption stages (ie, patients can make an online appointment with one specialism but not for another).

Another principal finding is that hospitals are struggling to adjust the portal to meet patients' needs in their continuous development process, although all seem to find this important. According to Greenhalgh et al [16], it is vital that intended users get the opportunity to refine the portal so that they will not drop out in the early adoption stage. Hospitals are using various ways to adjust to the patients' needs in ongoing portal development, without knowing which one is most effective in which phase and for what purpose. It would be interesting to do more research on how patients can participate in portal development, including efforts to stimulate adoption.

Nonusers of patient portals could be studied further. Previous studies show that nonusers have various reasons for not adopting the portal [36,37], including a preference to speak directly to their HCP, the level of their communication skills [36], and their concern for privacy and information security [36]. Such studies would show that hospitals are taking nonusers' concerns seriously and, at the same time, could produce insights valuable to exploring whether and how the patient portal could be made useful to them.

A remarkable finding is that portal suppliers do not include patients in their development process. The suppliers see it as their clients' responsibility to give voice to patients' wishes, but the focus group discussion did make them rethink this. This means that hospitals must explain to the supplier how they should make the portal more user-friendly for patients. Because of the variation in hospital context and portal suppliers, this could explain the disappointing adoption by patients.

Limitations

Our study has four limitations. The first is that the adoption processes in the hospitals we studied might be somewhat unusual due to the financial incentives of the national VIPP program. Conducting similar research in other countries would, therefore, be interesting and could also teach us more about the contextual, including cultural, factors that influence hospitals' efforts to stimulate adoption.

The second limitation is the way we recruited hospital participants. Using our own research networks may have biased our sample. However, our recruiting process resulted in a good variation in the mix of included hospitals.

The third limitation is that we only included teaching and general hospitals, given that academic hospitals follow another implementation program. Also, during the study period they were not participating in the VIPP program and, therefore, could not be compared. However, the inclusion criteria context of the studied hospitals varied greatly to include different kinds of hospitals and patients.

Last, this descriptive study shows the efforts that some Dutch hospitals have made to stimulate adoption of a patient portal. We did not study whether the undertaken efforts led to an actual increase of the adoption of patient portals. A further study on the effectiveness of these efforts is recommended.

Comparison With Prior Work

In recent years, many theoretical models on the adoption of information technology have been developed [9,10]. These

models show which variables are important for the adoption of a technology; for example, *perceived ease of use*, defined in the Technology Acceptance Model (TAM) as "the degree to which the person believes that using the particular system would be free of effort" [12]. However, these models are not explanatory and do not provide the know-how to stimulate patient portal adoption [9,10]. Consequently, we suggest future research should not focus on models, including new ones, but should deal with actionable knowledge for practice [38]. Action research can be used to study the adoption process and the embedding of patient portals in daily practice [39].

Communication experts support the hospitals' choice to use a cross-media promotion to inform patients and HCPs about the patient portal [40]. Explaining the benefits of using a patient portal is especially important for promotional messages [41]. However, open dialogue among HCPs and project leaders and staff is also vital because it illuminates the professionals' perspectives on portal development [41]. Earlier research shows that ignoring doubts while trying to convince others to use a technology may produce negative energy and a reluctance to use the portal [41,42]. Further research is required to find the most effective hospital communication strategies for encouraging patient portal adoption for patients and HCPs.

Our study showed that embedding a patient portal in the daily routine of HCPs requires changing their work processes. Earlier studies suggest that hospitals need to make extra time available to HCPs so that they can change and learn new work processes [41]. Research shows that portal use by patients may increase when HCPs are active on the patient portal and it is embedded in their work processes [43]. Research suggests training can benefit the adoption process [41], yet only some hospitals organize courses that explain how the portal works. As a result, some HCPs lack familiarity with portal functionalities [42].

The literature reports several ways of using feedback to adjust the portal to patients' needs in the continuous development of patient portals. These studies could help hospitals struggling with this. It is important to include patients at the beginning of ongoing development of patient portals [1,44,45]. Examples include co-design, where patients help identify the project based on personal experiences in collaborating with the clinician [45], and participatory stakeholder co-design, where patients and clinicians are equal stakeholders in the whole project [45]. Vulnerable patient groups, such as disadvantaged older adults, should be given special attention in the process of cocreation and user testing [7]. This is an important issue for future research.

Conclusions

Patient portal adoption processes are not just about implementing the technology. They require human interaction in a multitude of ways. Our study reveals three key findings for the adoption process: (1) informing patients and HCPs about the portal, (2) embedding it in the daily routine of HCPs, and (3) adjusting it to patients' needs in the continuous development of the portal. Our paper provides rich insights into the complexity of the adoption process and gives examples of efforts to stimulate the adoption of patient portals. Our findings help to translate the relatively abstract factors mentioned in the theoretical models to the everyday pragmatics of eHealth projects in hospitals.

Acknowledgments

The study was commissioned by the Dutch Hospital Association. We would like to thank our participants and the members of our scientific group. We are grateful for the helpful comments and insightful suggestions by our colleagues, especially Marcello Aspria and Louis Ter Meer.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Topic list for interviews.
[DOCX File , 22 KB - jmir_v22i5e16921_app1.docx]

References

1. Van Gemert-Pijnen L, Kelders SM, Kip H, Sanderman R, editors. eHealth Research, Theory and Development: A Multi-Disciplinary Approach. New York, NY: Routledge; 2018.
2. Otte-Trojel T, de Bont A, Rundall TG, van de Klundert J. What do we know about developing patient portals? A systematic literature review. J Am Med Inform Assoc 2016 Apr;23(e1):e162-e168 [FREE Full text] [doi: 10.1093/jamia/ocv114] [Medline: 26335985]
3. Al-Ramani M, Noteboom C. A systematic analysis of patient portals adoption, acceptance and usage: The trajectory for triple aim? In: Proceedings of the 51st Hawaii International Conference on System Sciences. 2018 Jan 03 Presented at: 51st Hawaii International Conference on System Sciences; January 2-6, 2018; Big Island, HI p. 852-861. [doi: 10.24251/hicss.2018.107]
4. Sieck CJ, Hefner JL, McAlearney AS. Improving the patient experience through patient portals: Insights from experienced portal users. Patient Exp J 2018 Nov 06;5(3):47-54. [doi: 10.35680/2372-0247.1269]

5. Kruse CS, Bolton K, Freriks G. The effect of patient portals on quality outcomes and its implications to meaningful use: A systematic review. J Med Internet Res 2015 Feb 10;17(2):e44 [FREE Full text] [doi: 10.2196/jmir.3171] [Medline: 25669240]
6. Grünloh C, Myreteg G, Cajander Å, Rexhepi H. "Why do they need to check me?" Patient participation through eHealth and the doctor-patient relationship: Qualitative study. J Med Internet Res 2018 Jan 15;20(1):e11 [FREE Full text] [doi: 10.2196/jmir.8444] [Medline: 29335237]
7. Wildenbos GA, Jaspers M, Peute L. The equity paradox: Older patients' participation in patient portal development. Int J Qual Health Care 2019 Dec 31;31(10):793-797. [doi: 10.1093/intqhc/mzy245] [Medline: 30576454]
8. Maguire D, Evans H, Honeyman M, Omojomolo D. Digital Change in Health and Social Care. London, UK: The King's Fund; Jun 2018.
9. Salahshour Rad M, Nilashi M, Mohamed Dahlan H. Information technology adoption: A review of the literature and classification. Univers Access Inf Soc 2017 Mar 30;17(2):361-390. [doi: 10.1007/s10209-017-0534-z]
10. Taherdoost H. A review of technology acceptance and adoption models and theories. Procedia Manuf 2018;22:960-967. [doi: 10.1016/j.promfg.2018.03.137]
11. Ajibade P. Technology Acceptance Model limitations and criticisms: Exploring the practical applications and use in technology-related studies, mixed-method, and qualitative researches. Libr Philos Pract 2018:1-13 [FREE Full text]
12. Marangunić N, Granić A. Technology Acceptance Model: A literature review from 1986 to 2013. Univers Access Inf Soc 2014 Feb 16;14(1):81-95. [doi: 10.1007/s10209-014-0348-1]
13. van Gemert-Pijnen L, Kelders SM, Beerlage-de Jong N, Oinas-Kukkonen H. Persuasive health technology. In: van Gemert-Pijnen L, Kelders SM, Kip H, Sanderman R, editors. eHealth Research, Theory and Development: A Multi-Disciplinary Approach. New York, NY: Routledge; 2018:228-246.
14. Pieterse M, Kip H, Cruz-Martínez RR. The complexitiy of eHealth implementation: A theoretical and practical perspective. In: Van Gemert-Pijnen L, Kelders SM, Kip H, Sanderman R, editors. eHealth Research, Theory and Development: A Multi-Disciplinary Approach. New York, NY: Routledge; 2018:247-270.
15. Benson T. Digital innovation evaluation: User perceptions of innovation readiness, digital confidence, innovation adoption, user experience and behaviour change. BMJ Health Care Inform 2019 Apr;26(1):1-6 [FREE Full text] [doi: 10.1136/bmjhci-2019-000018] [Medline: 31039121]
16. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: Systematic review and recommendations. Milbank Q 2004;82(4):581-629 [FREE Full text] [doi: 10.1111/j.0887-378X.2004.00325.x] [Medline: 15595944]
17. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
18. Rogers EM. Diffusion Of Innovations. 4th edition. New York, NY: The Free Press; Jul 06, 2010.
19. Williams MD, Rana NP, Dwivedi YK. The unified theory of acceptance and use of technology (UTAUT): A literature review. J Enterp Inf Manag 2015 Apr 13;28(3):443-488. [doi: 10.1108/JEIM-09-2014-0088]
20. Logue MD, Effken JA. Modeling factors that influence personal health records adoption. Comput Inform Nurs 2012 Jul;30(7):354-362. [doi: 10.1097/NXN.0b013e3182510717] [Medline: 22525046]
21. 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]
22. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Analysing the role of complexity in explaining the fortunes of technology programmes: Empirical application of the NASSS framework. BMC Med 2018 May 14;16(1):66 [FREE Full text] [doi: 10.1186/s12916-018-1050-6] [Medline: 29754584]
23. Triberti S, Kelders S, Gaggioli A. User engagement. In: Van Gemert-Pijnen L, Kelders SM, Kip H, Sanderman R, editors. eHealth Research, Theory and Development: A Multi-Disciplinary Approach. New York, NY: Routledge; 2018:271-289.
24. Wilkinson S. Focus group methodology: A review. Int J Soc Res Methodol 2014 Jun 10;1(3):181-203. [doi: 10.1080/13645579.1998.10846874]
25. Smithson J. Focus groups. In: Alasuurtari B, Bickman L, Brannen J, editors. The Sage Handbook of Social Research Methods. London, UK: SAGE Publications; 2008:357-370.
26. Frey JH, Fontana A. The group interview in social research. Soc Sci J 2019 Dec 09;28(2):175-187. [doi: 10.1016/0362-3319(91)90003-m]
27. VIPP. 2020. Wat is VIPP (What is VIPP) [article in Dutch] URL: https://www.vipp-programma.nl/over-vipp/wat-is-vipp [accessed 2020-02-11]
28. VIPP. Handboek VIPP Eindtoets (VIPP Handbook Final Test) [document in Dutch]. The Hague, the Netherlands: Ministerie van Volksgezondheid Welzijn en Sport (Ministry of Health, Well-Being and Sport); 2016 Dec 27. URL: https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/rapporten/2016/12/27/handboek-vipp-eindtoets/Handboek+VIPP+eindtoets+ter+publicatie2.pdf [accessed 2020-04-19]

29. Zoom.: Zoom Video Communications; 2020. URL: https://zoom.us/ [accessed 2020-02-11]
30. Morgan DL. Focus group interviewing. In: Gubrium JF, Holstein JA, editors. Handbook of Interview Research: Context and Method. London, UK: SAGE Publications; 2001:141-159.
31. Gibbs A. Focus groups and group interviews. In: Arthur J, Waring M, Coe R, Hedges LV, editors. Research Methods and Methodologies in Education. London, UK: SAGE Publications; 2012:186-192.
32. McLafferty I. Focus group interviews as a data collecting strategy. J Adv Nurs 2004 Oct;48(2):187-194. [doi: 10.1111/j.1365-2648.2004.03186.x] [Medline: 15369499]
33. Gaskell G. Individual and group interviewing. In: Bauer MW, Gaskell G, editors. Qualitative Researching with Text, Image and Sound: A Practical Handbook. London, UK: SAGE Publications; 2000:38-56.
34. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
35. Mortelmans D. Handboek Kwalitatieve Onderzoeksmethoden (Handbook of Qualitative Research Methods). 4th edition [book in Dutch]. Leuven, Belgium: Acco; 2013.
36. Anthony DL, Campos-Castillo C, Lim PS. Who isn't using patient portals and why? Evidence and implications from a national sample of US adults. Health Aff (Millwood) 2018 Dec;37(12):1948-1954. [doi: 10.1377/hlthaff.2018.05117] [Medline: 30633673]
37. Wyatt S. Non-users also matter: The construction of users and non-users of the Internet. In: Oudshoorn N, Pinch T, editors. How Users Matter: The Co-construction of Users and Technology. Cambridge, MA: The MIT Press; 2003:67-79.
38. Argyris C. Actionable knowledge: Design causality in the service of consequential theory. J Appl Behav Sci 2016 Jul 26;32(4):390-406. [doi: 10.1177/0021886396324004]
39. Coghlan D. Doing Action Research in Your Own Organization. 5th edition. London, UK: SAGE Publications; 2019.
40. Leppäniemi M, Karjaluoto H. Mobile marketing: From marketing strategy to mobile marketing campaign implementation. Int J Mob Mark 2008 Jun;3(1):50-61 [FREE Full text]
41. Kujala S, Hörhammer I, Kaipio J, Heponiemi T. Health professionals' expectations of a national patient portal for self-management. Int J Med Inform 2018 Sep;117:82-87 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.06.005] [Medline: 30032968]
42. Powell K, Myers CR. Electronic patient portals: Patient and provider perceptions. Online J Nurs Inform 2018 Feb 27;22(1):1 [FREE Full text]
43. Hoogenbosch B, Postma J, de Man-van Ginkel JM, Tiemessen NA, van Delden JJ, van Os-Medendorp H. Use and the users of a patient portal: Cross-sectional study. J Med Internet Res 2018 Sep 17;20(9):e262 [FREE Full text] [doi: 10.2196/jmir.9418] [Medline: 30224334]
44. Ryan BL, Brown JB, Terry A, Cejic S, Stewart M, Thind A. Implementing and using a patient portal: A qualitative exploration of patient and provider perspectives on engaging patients. J Innov Health Inform 2016 Jul 04;23(2):848 [FREE Full text] [doi: 10.14236/jhi.v23i2.848] [Medline: 27869582]
45. Kildea J, Battista J, Cabral B, Hendren L, Herrera D, Hijal T, et al. Design and development of a person-centered patient portal using participatory stakeholder co-design. J Med Internet Res 2019 Feb 11;21(2):e11371 [FREE Full text] [doi: 10.2196/11371] [Medline: 30741643]

Abbreviations

CMIO: Chief Medical Information Officer
CNIO: Chief Nursing Information Officer
DigiD: Dutch national identity authentication method
DRG: diagnosis-related group
GDPR: General Data Protection Regulation
eHealth: electronic health
EHR: electronic health record
Erasmus MC: Erasmus Medical Center
HCP: health care professional
NASSS: nonadoption, abandonment, scale-up, spread, and sustainability
NICTIZ: Nationaal ICT Instituut in de Zorg
PHR: personal health record
TAM: Technology Acceptance Model
VIPP: Versnellingsprogramma Informatie-Uitwisseling Patiënt en Professional

Edited by G Eysenbach; submitted 05.11.19; peer-reviewed by S Wiig, S Choi, K Goniewicz, G Jambaulikar, B van Lettow; comments to author 06.01.20; revised version received 02.03.20; accepted 05.03.20; published 19.05.20.

Please cite as:
Hulter P, Pluut B, Leenen-Brinkhuis C, de Mul M, Ahaus K, Weggelaar-Jansen AM
Adopting Patient Portals in Hospitals: Qualitative Study
J Med Internet Res 2020;22(5):e16921
URL: http://www.jmir.org/2020/5/e16921/
doi:10.2196/16921
PMID:32427110

Original Paper

Effects of a Web-Based Patient Portal on Patient Satisfaction and Missed Appointment Rates: Survey Study

Timothy A D Graham[1,2], MD, MSc, CPHIMS-CA; Samina Ali[1,3], MDCM; Melita Avdagovska[1], MA; Mark Ballermann[2,4], MSc, PhD

[1]School of Public Health, University of Alberta, Edmonton, AB, Canada

[2]Alberta Health Services, Edmonton, AB, Canada

[3]Women and Children's Health Research Institute, University of Alberta, Edmonton, AB, Canada

[4]Department of Critical Care Medicine, University of Alberta, Edmonton, AB, Canada

Corresponding Author:
Timothy A D Graham, MD, MSc, CPHIMS-CA
School of Public Health
University of Alberta
116 St & 85 Ave
Edmonton, AB,
Canada
Phone: 1 7803423464
Email: tim.graham@ahs.ca

Abstract

Background: Although electronic medical record (EMR)-tethered patient portals are common in other countries, they are still emerging in Canada.

Objective: We aimed to report user satisfaction and the effects of a patient portal on medical appointment attendance in a Canadian cohort of patients within our publicly funded health care system.

Methods: Two surveys were deployed, via email, at 2 weeks and 6 months following the first recorded patient portal access. Database audits of visit attendance were used to supplement and cross reference survey data.

Results: Between January 2016 and July 2018, 4296 patients accessed the patient portal. During the study, 28% (957/3421) consented patient portal users responded to one or more semistructured electronic surveys. Of respondents, 93% (891/957) reported that the patient portal was easy to use, 51% (492/975) reported it saved time when scheduling an appointment, and 40% (382/957) reported that they had to repeat themselves less during appointments. Respondents reported patient portal–related changes in health system use, with 48% (462/957) reporting avoiding a clinic visit and 2.7% (26/957) avoiding an emergency department visit. Across 19,968 visits in clinics where the patient portal was introduced, missed appointments were recorded in 9.5% (858/9021) of non–patient portal user visits, compared with 4.5% (493/9021) for patient portal users, representing a 53% relative reduction in no-show rates.

Conclusions: Early experience with an EMR-tethered patient portal showed strong reports of positive patient experience, a self-reported decrease in health system use, and a measured decrease in missed appointment rates. Implications on the expanded use of patient portals requires more quantitative and qualitative study in Canada.

(J Med Internet Res 2020;22(5):e17955) doi:10.2196/17955

KEYWORDS

patient portal; patient satisfaction

Introduction

Canadians desire electronic access to their health information and have an expectation of managing their health system interactions digitally [1]. Increasingly, that access is provided through electronic medical record (EMR)-tethered and standalone web-based patient portals [2-12]. Patient portals can deliver digital access to secure messaging, health information (eg, test results, lists of medications and allergies), scheduling functions, and self-management of health issues such as weight, blood glucose, and blood pressure. Early studies of patient portals reported patient benefits including enhanced satisfaction,

better relationships with their care providers, more efficient medication refills, and improved understanding of their health information [3,4,11,13-16]. Health system benefits include fewer missed clinic appointments (no-shows), lower postage costs related to decreased mailing of paper records to patients, and less time answering phone calls related to appointments [3,4,11,13-16]. A recent systematic review of patient portals concluded that there is an urgent need for more outcomes data and reporting on organizational and provider context, as well as implementation processes [17].

Fully interactive patient portals that share detailed health information, messaging, and scheduling capability are relatively rare in Canada, and mostly allow patients limited interaction with a single health care facility or clinic [2,18-22]. A number of provinces provide standalone portals that aggregate laboratory and diagnostic imaging results, but few employ patient portals that allow self-scheduling or health management across a large population. Alberta Health Services (AHS) had a limited production rollout of a patient portal tethered to an EMR in the Edmonton zone from 2016-2019. The objectives of our study were to examine the effects of this tethered patient portal on patient satisfaction and health system usage.

Methods

Design

For this study, we employed surveys and system audit logs. This combination of longitudinal semistructured user surveys and administrative data audit was used to identify user-reported outcomes triangulated with possible clinical outcomes of interest in a cohort of users attending clinics allowing access to a patient portal. The University of Alberta's Health Research Ethics Board (Pro00066287) approved the study.

Patient Portal Development and Adoption

The patient portal that we studied is tethered to a shared EMR known as eCLINICIAN (an AHS branding of EpicCare Ambulatory 2014 from Epic Systems), used at the time by approximately 1110 physicians, residents, and students, and approximately 6000 nurses, allied health professionals, and other staff. The adaptation of this patient portal to the AHS context took about six months, the bulk of which was spent achieving consensus (including patient, clinical, administrative, legislative, and regulatory college input) on policy decisions and functions (Textbox 1). Consensus was reached to release most high-value data in real time (ie, those data required to manage active or chronic disease), with higher stakes information (ie, results of hepatitis or biopsies) delayed by ten days (Textbox 2).

Textbox 1. Patient portal features.

- Secure bidirectional messaging between patient and care team
- Display medical information (medications, problems, past history, allergies, immunizations)
- Create, select, view, and modify scheduled appointments (varies by clinic)
- View appointments and receive reminders sent to patient mobile devices
- Complete previsit questionnaires and patient-entered flowsheets
- Enter and track home data (eg, blood pressure, glucose)
- iOS and Android mobile apps
- Integration with Alberta Personal Health Portal
- Email signup process to reduce initial 2-factor login issues for patients

Textbox 2. Results release framework.

Results released in real time (hourly)

- Hematology (complete blood count, white blood cell count), electrolytes, renal function, liver function, lipids, diabetes monitoring (glucose, HbA_{1C}), coagulation profile (international normalized ratio, prothrombin time test), cardiac tests (troponin, B-type natriuretic peptide), sexually transmitted infection results, bacterial cultures

Results released after 10-day delay (or earlier if provider does so manually)

- Pathology reports, diagnostic imaging reports, genetic testing (newborn screening, specific diagnostic testing (eg, for angioedema, protein C deficiency, carrier testing, prenatal testing, preimplementation testing), cytogenetics, histocompatibility, serology, molecular diagnostics, HIV serology, hepatitis C serology, cancer-related markers (eg, prostate-specific antigen, carcinoembryonic antigen)

Study Setting and Population

AHS is a provincial health system responsible for the delivery of inpatient and ambulatory care services. Edmonton is the capital city of Alberta, with a population of approximately one million. In January 2016, an EMR-tethered patient portal was introduced in the Edmonton zone, and was evaluated between January 2016 and July 2018. During this phase, the patient portal was introduced in one family practice clinic and four specialty clinics (rheumatology, inflammatory bowel disease, multiple

sclerosis, and diabetes). These specialty clinics are part of a large academic health campus, accept patients initially via referral for specialty care, then follow them for their chronic disease management. Inclusion criteria for the survey study were the following: patients who had (1) visited 1 of the 5 aforementioned clinics between January 2016 and July 2018, and (2) signed up for patient portal access. A total of 3421 patients met inclusion criteria for the survey during the study period.

Patient Survey Instrument

A novel 30-question survey instrument (Multimedia Appendix 1) was developed by members of our research team using a modified Delphi [23] approach. Four overarching themes were identified: (1) satisfaction with the patient portal, (2) utility of the patient portal, (3) impact of the patient portal, and (4) demographic characteristics. Proposed questions were refined by the research team and then pretested with 6 members of our patient portal working group. After further refinement, which included changes and clarification of wording and question structure, the survey tool was piloted and sensibility-tested with a small group of patients, leading to a final survey with 5 (satisfaction, utility) to 12 (impact) questions per theme. The survey required about 10-15 minutes to complete.

During the sign-up process for the patient portal, patients consented to be contacted via their email address, which was routinely recorded as part of the sign-up process. Patient portal users were emailed a voluntary link to the survey which contained Likert scales, multiple-choice questions and free-text questions. The free-text answers were collated and thematically coded by two authors (TG, SA). Coding discrepancies were resolved by consensus. Surveys were administered at 2-week and 6-month time points relative to each participant's first successful access to the patient portal, as this was felt to represent an adequate period of time for participants to access and become comfortable with using the patient portal. Respondents were made aware that their responses would not impact their clinical care, that no incentives were provided, and that their responses would only be studied in collated form. Respondents were permitted to skip any question they wished.

Appointment Attendance Records

To determine the no-show rate, the numerator of scheduled appointments recorded as no-show was divided by the denominator of total scheduled appointments. No-show rates were calculated for users with an active patient portal account at the time of the visit and compared to rates for users attending the same clinics without a patient portal account at the time of the visit. Audit log analysis included attended and no-show visit rates, basic patient demographic data, and the patient's contact email address.

Statistical Analysis

Categorical data (eg, sex, age) were summarized with frequency distributions. Respondent demographic characteristics were compared to the wider user population with respect to sex and age. No-show rates were compared for patient visits where a patient portal account was active at the time and patient visits in the same clinic with no active patient portal account between November 2015 (just before the patient portal was introduced) and July 2018. Each respondent's final answer across the 3 time points to questions of a more summative nature were included for statistical testing. Statistical differences were detected at a *P* value of less than .05. Chi-square testing for each of these comparisons was conducted using R statistical software (version 3.4.3, R Core Team, R Foundation for Statistical Computing). We included users who enrolled to use the system between May 2016 and May 2018. An additional 875 new users signed up for patient portal access between May and July of 2018. To avoid biasing the estimate of the proportions of users toward the early users, these users were included in the user population but not in survey analysis.

Results

Demographic Characteristics

A total of 5629 electronic surveys were transmitted to 3421 patients between January 2016 and July 2018. The final administration of the survey was completed in May 2018 to allow users 2 months to respond to the survey. Overall, 957 patients responded 1916 times, and 28% (957/3421) of users receiving an invitation to participate replied to at least one survey. In addition to accessing their own health information, patient portal users reported using the patient portal for their children (36/957, 4%) or their spouse (131/957, 16%). Of 174,298 login sessions, 70% (122,648/174,298) were from a computer browser, 20% (34,205/174,298) were from an iPhone, and 10% (17,445) were from an Android device.

Table 1 describes user and respondent demographics and comfort with computer use; the majority of respondents (834/957, 87%) reported being either comfortable or completely comfortable using computers).

Table 1. Patient portal user and respondent demographics.

Characteristics	Respondent frequency, n (%), (N=957)	User frequency, n (%), (N=4296)	*P* value
Sex			>.99
Female	537 (56.1)	2347 (59.3)	
Male	378 (39.5)	1749 (40.7)	
No answer	43 (4.3)	N/A[a]	
Age (years)			.22
<18	0 (0)	73 (1.7)	
18-29	64 (6.7)	719 (16.7)	
30-39	111 (11.6)	870 (20.3)	
40-49	132 (13.8)	742 (17.3)	
50-59	249 (26.0)	893 (20.8)	
60-69	200 (20.9)	630 (14.7)	
≥70	105 (11.0)	369 (8.6)	
No answer	96 (10.0)	N/A	
Comfort with computers			
Completely comfortable	297 (31.0)	N/A	
Comfortable	537 (56.1)	N/A	
Neutral	297 (31.0)	N/A	
Uncomfortable	80 (8.4)	N/A	
Completely uncomfortable	2 (0.2)	N/A	
No answer	34 (3.6)	N/A	

[a]N/A: Not applicable.

Satisfaction, Impact, and Utility

Patient portal users reported a high degree of usability and general satisfaction: 93% (891/957) of respondents felt the patient portal was easy to use, 83% (794/957) said it made communication more convenient, and 75% (716/957) indicated it saved time when scheduling an appointment.

Among survey respondents, 48% (460/957) stated the patient portal had helped them avoid a clinic visit. In February 2016, we added an additional question regarding emergency department or urgent care visit avoidance, and of users who answered this, 14% (26/188) answered that the patient portal allowed them to avoid such a visit.

There were 4 free-text questions available based on the themes of requested improvements, preferred features, how the patient portal helped avoid a health care visit (if applicable), and self-reported cost savings related to the patient portal (Table 2). Over two-thirds of respondents identified that access to laboratory and diagnostic imaging results was the most preferred feature of the patient portal. The most frequently reported concerns were related to a cumbersome sign-in process and a desire for more explanation of results. It should be noted that the patient portal was launched with an authentication method that was acknowledged to be overly complex at the time of implementation but was required for local technical and security reasons. Less common requests included a wider availability of data, including physician notes on the chart, an explanation of the meaning of the results, and the timeliness of the reports in the system.

Table 2. Respondents' view on the patient portal.

Participant responses[a]	Frequency, n (%)
Requested improvements (n=558)	
Less cumbersome login authentication	144 (25.8)
More data types and wider spread availability	143 (25.6)
Like it/love it in current state	94 (16.8)
Better instructions/explanation of results	47 (8.4)
Reports delayed or incomplete	43 (7.7)
Other[b]	108 (19.4)
Most preferred features (n=173)	
Lab and diagnostic imaging results	119 (68.8)
Feeling of health empowerment	23 (13.3)
Scheduling features	21 (12.1)
Communication features	14 (8.1)
Other[c]	4 (2.3)
Factors resulting in self-reported avoidance of an emergency department/urgent care center visit (n=129)	
Access to results meant a visit could be avoided	76 (58.9)
Chronic disease–specific management	22 (17.1)
Question answered via secure message[d]	21 (16.3)
Other[e]	50 (38.8)
Reported cost savings by avoiding a health care visit (n=112)	
General convenience and less waiting in clinic	43 (38.4)
Not paying for parking and gas	25 (22.3)
Overall improved efficiency for the health system	19 (17.0)
Not paying for meals/hotels	10 (8.9)
Other[f]	2 (1.8)

[a]There was more than one answer permitted per category.

[b]Other included general usability and results management (n=41), adding and updating health information (n=20), view diagnostic imaging images (n=14), secure messaging and notifications (n=14), patients able to upload pictures and documents (n=11), and better phone/tablet experience (n=8).

[c]Other included access to health information decreases anxiety (n=7), can view medications (n=7), proxy access to family member's chart (n=4), and previsit questionnaires (n=1).

[d]Secure messaging was done through MyChart and answered a question from the user.

[e]Other included visit more efficient because informed (n=9), health empowerment (n=8), and preferred to get results in person (n=3).

[f]Other included not paying for meals/hotels (n=10), not missing work (n=7), and not paying for childcare (n=2).

Health System Usage

We studied the proportion of patients who had a no-show to an appointment, as it related to patient portal user status. From January 2016 to July 2018, there were 19,968 no-shows across the 5 clinics that participated. Of these visits, 9021 did not have an active patient portal account at the time of the visit. The number of no-shows amongst these visits for patients without a patient portal was 858 (9.5%, range: 3.7%-16.6%). Of the remaining 10,947 visits for patients with an active patient portal account, 493 (4.5%) no-show visits were recorded (range: 3.1%-10.4%). No-show rates for patients with no patient portal were 11.9% (386/3257) for the specialty clinics and 8.2% (472/5764) for the family practice clinic, whereas the rates for patients with patient portal access were 4.9% (194/3943) and 4.3% (299/7004), respectively. This represented a 53% relative reduction in the no-show rate overall (P<.001).

Discussion

Overview

In our study, patients were surveyed about their perceptions of an EMR-tethered patient portal. Patients had high general satisfaction, with over 90% reporting that it was easy to use, and almost half reporting that it saved them a medical visit. Respondents reported that access to results and advanced features, such as messaging and scheduling, improved

communication with their care providers. Additionally, self-reported reductions in health system use implied improved utilization of scarce health system resources. Finally, objective data from clinic visits demonstrated an over 50% relative reduction in no-show rates.

We decided to focus on patient-centered indicators of patient portal utility rather than tracking health outcomes in our study. As with any technology, the benefits will accrue only if the technology itself is working and is being used as expected [24]. Although it is assumed that just having access to health information is of benefit, there is little evidence that access to lab tests will improve health outcomes. However, we believe that more advanced features, such as scheduling and secure messaging, in combination with access to test results and clinical notes are more likely to result in long-term health system benefits. In 2014, Goldzweig et al [17] performed a systematic review and found that there were a limited number of studies about patient portals, with heterogeneous designs making it difficult to draw strong conclusions about a new technology. Additionally, they found that improved outcomes (eg, for chronic disease such as diabetes, hypertension, and depression) tended to involve portal use in conjunction with case management [17]. They also found mixed data about the effect of portals on health care utilization and efficiency. Based on the success of the patient portal in our study, planning is underway to expand patient portal use across all AHS clinics. This expansion is targeted at the notion that patients having access to their own information in conjunction with the ability to interact with the health system digitally will contribute to increasing health system efficiency, including reductions in no-show rates and clinic and emergency department (ED) visits, increased patient satisfaction and empowerment and, in the longer term, improved health outcomes. The latter assumes that the patient portal is not only widely available, but actively being used as intended.

The 53% relative reduction in the no-show rate seen in patient portal users in the 5 pilot clinics may be widely relevant. Across the Edmonton zone of AHS, about 200,000 appointments are scheduled monthly**.** Of these**,** about 40,000 appointments are rescheduled, and 2000 appointments are considered no-shows. Patients who miss appointments often use emergency departments as sources of both primary and chronic care, driving up costs and straining hospital systems [23-25]. Missed appointments can compromise continuity and quality of care for both the patients who no-show and others who would have been scheduled in those appointment slots [25]. If the decreased no-show rates scale up to a wider set of users, patient portals may have significant positive impact on wait times and clinic efficiency. Access to scheduling functions in a patient portal may also be associated with decreased urgent care visits, although studies are not consistent [26,27]. In the current work, 2.7% of respondents reported being able to avoid one or more ED or urgent care visits, and 48% reported avoiding a clinic visit due to being able to communicate electronically. ED overcrowding is an ongoing and seemingly intractable problem; any strategy to safely decant these pressures in the Canadian context warrants further exploration.

In the near future, patients will expect web-based and mobile app–based access to their health data, and the ability to manage their health and interact with the health system digitally. In some jurisdictions, this access is not only related to tests and scheduling, but extends to transparent access into the complete medical record including progress notes [28]. Future studies will include interviews of care providers and patients to get a deeper understanding of the pros and cons of an EMR-tethered patient portal, its effect on clinic workflows, and the potential savings, whether that be in terms of time saved not waiting on hold during a phone call, or travel costs saved by replacing an inpatient visit with a secure message.

Limitations

Our study evaluated a single patient portal in a large group of patients attending 1 of 5 clinics, in a single city, where the clinic leads were keen to implement it. It is not clear whether the results would be broadly generalizable. However, the study population did include both family medicine and specialty clinics, representing some diversity of needs and practice. Self-reported data is subject to recall bias, which we attempted to limit by sampling at different times after a user's first successful patient portal access. The correlation between patient portal use and no-show rates may be partly the result of more diligent or engaged patients at the clinics participating in the study. Although we were able to compare survey respondent versus non-respondent demographic characteristics for all patients that had patient portal access, we did not have the ability to do the same for those with patient portal access versus those without. The analysis includes visits from non–patient portal users who eventually obtained patient portal access, which may underreport the effect size. The respondent population reported a high degree of comfort with computers, which may not hold true elsewhere, although it may reflect increasing competence with technology in general.

Conclusions

Early experience with an EMR-tethered patient portal showed decreased no-show rates for appointments, high patient satisfaction, and self-reported changes in health system use. Implications on the expanded use of patient portals requires further study in Canada.

Acknowledgments

This study received an *EMR Innovation Grant* from the University of Alberta (September 2016 to September 2018).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey Instrument.
[PDF File (Adobe PDF File), 99 KB - jmir_v22i5e17955_app1.pdf]

References

1. Canada Health Infoway. ACCESS Health URL: https://www.infoway-inforoute.ca/en/solutions/access-health [accessed 2020-04-19]
2. my ehealth. URL: http://www.myehealth.ca [accessed 2018-12-21]
3. Chen C, Garrido T, Chock D, Okawa G, Liang L. The Kaiser Permanente Electronic Health Record: transforming and streamlining modalities of care. Health Affairs 2009 Apr;28(2):323-333 [FREE Full text] [doi: 10.1377/hlthaff.28.2.323] [Medline: 19275987]
4. Nazi K. Veterans' voices: use of the American Customer Satisfaction Index (ACSI) Survey to identify My HealtheVet personal health record users' characteristics, needs, and preferences. J Am Med Inform Assoc 2010;17(2):203-211 [FREE Full text] [doi: 10.1136/jamia.2009.000240] [Medline: 20190065]
5. Rigby M, Georgiou A, Hyppönen H, Ammenwerth E, Keizer ND, Magrabi F, et al. Patient Portals as a Means of Information and Communication Technology Support to Patient-Centric Care Coordination – the Missing Evidence and the Challenges of Evaluation. Yearb Med Inform 2018 Mar 10;24(01):148-159. [doi: 10.15265/iy-2015-007]
6. Otte-Trojel T, de Bont A, Rundall TG, van de Klundert J. What do we know about developing patient portals? a systematic literature review. J Am Med Inform Assoc 2016 Apr 02;23(e1):e162-e168 [FREE Full text] [doi: 10.1093/jamia/ocv114] [Medline: 26335985]
7. Ammenwerth E, Schnell-Inderst P, Hoerbst A. The impact of electronic patient portals on patient care: a systematic review of controlled trials. J Med Internet Res 2012;14(6):e162 [FREE Full text] [doi: 10.2196/jmir.2238] [Medline: 23183044]
8. Otte-Trojel T, de Bont A, Rundall TG, van de Klundert J. How outcomes are achieved through patient portals: a realist review. J Am Med Inform Assoc 2014 Jul;21(4):751-757. [doi: 10.1136/amiajnl-2013-002501] [Medline: 24503882]
9. Irizarry T, DeVito DA, Curran CR. Patient Portals and Patient Engagement: A State of the Science Review. J Med Internet Res 2015;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
10. Risling T, Martinez J, Young J, Thorp-Froslie N. Evaluating Patient Empowerment in Association With eHealth Technology: Scoping Review. J Med Internet Res 2017 Sep 29;19(9):e329 [FREE Full text] [doi: 10.2196/jmir.7809] [Medline: 28963090]
11. Wells S, Rozenblum R, Park A, Dunn M, Bates DW. Organizational strategies for promoting patient and provider uptake of personal health records. J Am Med Inform Assoc 2015 Jan;22(1):213-222 [FREE Full text] [doi: 10.1136/amiajnl-2014-003055] [Medline: 25326601]
12. Brudnicki S, Dalziel S. The BMJ Opinion. Collaborating with patients to provide online access to health records URL: http://blogs.bmj.com/bmj/2018/04/25/collaborating-with-patients-to-provide-online-access-to-health-records/ [accessed 2018-08-14]
13. Greenhalgh T, Stramer K, Bratan T, Byrne E, Russell J, Hinder S, et al. The devil's in the detail: final report of the independent evaluation of the Summary Care Record and HealthSpace programmes. London: University College London; 2010.
14. Serrato CA, Retecki S, Schmidt DE. MyChart-A New Mode of Care Delivery: 2005 Personal Health Link Research Report. Perm J 2007;11(2):14-20 [FREE Full text] [Medline: 21461088]
15. Mansour D. Better Health Partnership. Patient Portals Internet URL: http://www.betterhealthpartnership.org/pdfs/lc_presentations/2012_03/PM-Patient-Portals_lr.pdf [accessed 2016-03-22]
16. Halamka JD, Mandl KD, Tang PC. Early Experiences with Personal Health Records. Journal of the American Medical Informatics Association 2008 Jan 01;15(1):1-7. [doi: 10.1197/jamia.m2562]
17. Goldzweig CL, Orshansky G, Paige NM, Towfigh AA, Haggstrom DA, Miake-Lye I, et al. Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. Ann Intern Med 2013 Nov 19;159(10):677-687. [doi: 10.7326/0003-4819-159-10-201311190-00006] [Medline: 24247673]
18. Webster PC. Canada's electronic health records initiative stalled by federal funding freeze. CMAJ 2010 Mar 23;182(5):E247-E248 [FREE Full text] [doi: 10.1503/cmaj.109-3183] [Medline: 20159901]
19. Green M. Canada Health Infoway. 2016. Nova Scotia Online Patient Portal Leading the Way Internet URL: https://www.infoway-inforoute.ca/en/what-we-do/blog/consumer-health/7140-nova-scotia-online-patient-portal-leading-the-way [accessed 2019-12-21]
20. eChart Manitoba. URL: http://www.echartmanitoba.ca [accessed 2018-12-21]
21. New Brunswick Department of Health. Health Portal URL: https://hps.gnb.ca/my.policy [accessed 2018-12-21]
22. Canadian Healthcare Technology. Quebec moves to a single province-wide EHR URL: http://www.canhealth.com/blog/quebec-moves-to-a-single-province-wide-ehr/ [accessed 2018-12-21]
23. Rowe G, Wright G. Expert Opinions in Forecasting: The Role of the Delphi Technique. Principles of Forecasting 2nd ed. Boston, MA: Springer US; 2001:125.
24. Price M, Lau F. The clinical adoption meta-model: a temporal meta-model describing the clinical adoption of health information systems. BMC Med Inform Decis Mak 2014 May 29;14:43 [FREE Full text] [doi: 10.1186/1472-6947-14-43] [Medline: 24884588]

25. Parker MM, Moffet HH, Schillinger D, Adler N, Fernandez A, Ciechanowski P, et al. Ethnic Differences in Appointment-Keeping and Implications for the Patient-Centered Medical Home-Findings from the Diabetes Study of Northern California (DISTANCE). Health Serv Res 2011 Oct 27;47(2):572-593. [doi: 10.1111/j.1475-6773.2011.01337.x]
26. Solberg LI, Maciosek MV, Sperl-Hillen JM, Crain AL, Engebretson KI, Asplin BR, et al. Does improved access to care affect utilization and costs for patients with chronic conditions? Am J Manag Care 2004 Oct;10(10):717-722 [FREE Full text] [Medline: 15521163]
27. Belardi F, Weir S, Craig F. A controlled trial of an advanced access appointment system in a residency family medicine center. Fam Med 2004 May;36(5):341-345 [FREE Full text] [Medline: 15129381]
28. Walker J, Darer JD, Elmore JG, Delbanco T. The road toward fully transparent medical records. N Engl J Med 2014 Jan 02;370(1):6-8. [doi: 10.1056/NEJMp1310132] [Medline: 24304001]

Abbreviations

AHS: Alberta Health Services
ED: emergency department
EMR: electronic medical record

Edited by G Eysenbach; submitted 23.01.20; peer-reviewed by S Khan, F Kaliyadan; comments to author 15.03.20; revised version received 23.03.20; accepted 23.03.20; published 19.05.20.

Please cite as:
Graham TAD, Ali S, Avdagovska M, Ballermann M
Effects of a Web-Based Patient Portal on Patient Satisfaction and Missed Appointment Rates: Survey Study
J Med Internet Res 2020;22(5):e17955
URL: http://www.jmir.org/2020/5/e17955/
doi:10.2196/17955
PMID:32427109

Original Paper

Tracing the Decisions That Shaped the Development of MyChart, an Electronic Patient Portal in Alberta, Canada: Historical Research Study

Melita Avdagovska[1], MA; Tania Stafinski[1], PhD; Mark Ballermann[2], PhD; Devidas Menon[1], PhD; Karin Olson[3], PhD, RN; Pauline Paul[3], PhD, RN

[1]University of Alberta, School of Public Health, Edmonton, AB, Canada

[2]University of Alberta, Faculty of Medicine and Dentistry, Edmonton, AB, Canada

[3]University of Alberta, Faculty of Nursing, Edmonton, AB, Canada

Corresponding Author:
Melita Avdagovska, MA
University of Alberta
School of Public Health
3021, Research Transition Facility
8308 114 Street
Edmonton, AB, T6G 2V2
Canada
Phone: 1 7809083334
Email: avdagovs@ualberta.ca

Abstract

Background: Understanding how health organizations decide on information technology (IT) investments is imperative to ensure successful implementation and adoption. There is a high rate of failure and a tendency to downplay the complexity of implementation progression. Alberta Health Services introduced a patient portal called MyChart. Although MyChart allows patients to view appointments and selected laboratory results and to communicate with their providers, its uptake varies.

Objective: The study aimed to examine the institutional decision-making processes that shaped the development and implementation of MyChart.

Methods: A historical study was conducted based on the 7-step framework, where one engages in a rigorous archival critical analysis (including internal and external criticism) of documents and analysis of interviews. We reviewed and analyzed 423 primary and secondary sources and interviewed 10 key decision makers.

Results: Supportive leadership, project management, focused scope, appropriate technology and vendor selection, and quick decision making were some of the facilitators that allowed for the growth of proof of concept. The planning and implementation stages did not depend much on the technology itself but on the various actors who influenced the implementation by exerting power. The main barriers were lack of awareness about the technology, proper training, buy-in from diverse system leaders, and centralized government decision making.

Conclusions: Organizational priorities and decision-making tactics influence IT investments, implementation, adoption, and outcomes. Future research could focus on improving the applicability of needs assessments and funding decisions to health care scenarios.

(J Med Internet Res 2020;22(5):e17505) doi:10.2196/17505

KEYWORDS

patient portals; information technology; decision making; investments

Introduction

Backgound

Aging populations, increased prevalence of chronic conditions, and consequent rising costs significantly challenge health care systems worldwide. One proposed solution to these challenges has been health information technologies (ITs) that empower patients to be partners in their care, support evidence-based, individualized care, and monitor population health [1]. Understanding how health organizations decide on IT investments is imperative to ensure successful implementation and adoption. There is a high rate of failure and a tendency to downplay the complexity of implementation progression.

The literature shows that the majority of health IT investments are struggling to achieve the hoped-for improvements in quality of care and economic benefits [2-5]. Furthermore, these technologies tend to run over budget because of inadequate preparation for the complexities of implementation [6,7]. There is a need to understand the implementation and adoption of such technologies through a life cycle approach for the technology rather than as a decision at a single point in time. Understanding the events and actors involved in each stage of the cycle provides for better future planning for the successful implementation of IT investments such as patient portals [8]. Alberta Health Services (AHS) introduced a patient portal called MyChart through a proof of concept (PoC).

Aim and Objectives

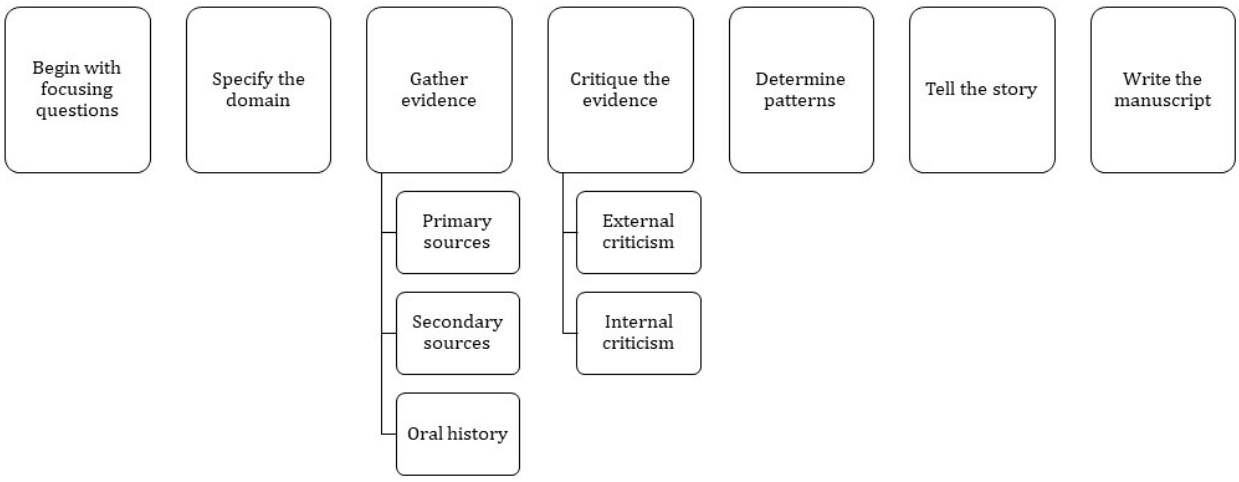
This study aimed to investigate and describe the process by which health IT, in this case, a patient portal, was introduced into the provincial health system of Alberta, Canada. The focus was on the process of decision making and the chronological timelines that led to the pilot of the patient portal, with an emphasis on the conceptualization, development, and implementation processes: need (why), process (how), decision makers (who), decision (what), setting and context (where), and timelines (when).

Methods

Study Design

A historical research approach was used to trace the history of the development and implementation of a patient portal in several clinics in Alberta [9-13]. Specifically, the 7-step methodology framework developed by Mason et al [9] for studying medical information systems was used, as shown in Figure 1. This entails a rigorous archival critical analysis (including internal and external criticism) of numerous documents (contracts, meeting agendas and minutes, training and marketing materials, reports, decision requests, etc), analysis of key informant interviews, and development of the narrative [9].

Figure 1. Mason et al (1997) 7-step framework.



Historical research includes the methodical collection and appraisal of data to recognize, understand, explain, illuminate, and accurately reconstruct past events, actions, and decisions [10-12]. Organizations may associate IT solutions with *awesome potential* and lose track of concerns and problems, resulting in repetitive regurgitation of ideas and being victims of IT fads and fashions. Mason et al [9] pioneered a framework to describe how a solution and its identified need may be adopted when it encounters the organizational context, including the connections and roles of change agents working to mitigate resistance. History allows for understanding and acknowledging what has worked and what has not worked previously [9]. It is important to outline that studying IT through the historical method is not about understanding the technology. It is about the connections and roles that impact how the technology is implemented and eventually used [9]. The framework developed by Mason et al [9] allows for an in-depth understanding of the organization' s current practices, how cultural and environmental conditions impacted the decisions, how the need and the problems were identified, shedding light on the resistance, the process of change, and the actors that led the change.

Study Setting

Alberta Health (AH) is the Alberta government department accountable for ensuring the delivery of health care services and setting and assessing compliance with policy and legislation

[14]. Health care service delivery is funded through AHS and primary care networks (PCNs). PCNs coordinate the delivery of primary health services [15,16]. AHS was established in 2008 and delivers care through 400 facilities throughout the province [17]. The focus of this historical study is the MyChart PoC that took place in Edmonton, Alberta.

Study Context

Planning for a patient portal in Alberta began in the early 2000s led by AH. However, it was not until 2016 that one was introduced into the province, with the launch of a PoC study of MyChart (AHS branding of EpicCare Ambulatory from Epic Systems, 2014 version) by AHS. MyChart was developed by Epic and customized to meet the needs of AHS, allowing patients to view appointments, medical test results, and medication therapies, and to communicate with their providers. It is connected to the central AHS electronic medical record (EMR), named eClinician.

This study received ethics approval from the Research Ethics Board at the University of Alberta (Study ID: MS1_Pro00072286).

Search Strategy

Information on the introduction of patient portals in Alberta was gathered by first identifying relevant sources. Primary sources, including oral histories (key informant interviews), are materials that provide firsthand accounts of the event of interest [18-20]. Oral histories are considered a primary source, as the interviews are *for the record* and tend to confirm the events outlined in written documents [21]. Secondary sources are reports, materials, books, or articles written on the topic of interest by people who were not directly involved [10,21]. They provide additional *depth and meaning to a topic* [22].

We developed a search strategy to identify any sources of the development and/or implementation of a patient portal in Alberta. As the implementation of patient portals was conducted internally, government archives were searched in February 2018, and it was identified that AHS and Alberta Health maintained the sources. Requests were submitted, and the AHS project leadership agreed to provide the documents. In addition, MA (lead researcher) contacted the provincial and AHS archive departments, but the archivists found no documents on this topic. It was expected that most of the written data sources (both primary and secondary) would be internal documents; thus, an agreement was signed with AHS for document access.

In addition, we conducted a Web-based search of academic electronic databases (PubMed [MEDLINE and non-MEDLINE] references and Cumulative Index of Nursing and Allied Health Literature). The search focused on the term *Alberta patient portal* with the intent of identifying articles that had any descriptions about the development of a patient portal in Alberta.

Sampling Procedure

Purposive sampling was used to recruit key informants from the AHS and AH. The names of possible participants were identified through meetings with AHS and Alberta Health representatives and a search of government directories. Both Alberta Health and AHS individuals were interviewed because of their role in planning for a patient portal. This sampling approach is common in historical studies, as the researcher requires information from individuals with firsthand knowledge of the topic under investigation. Furthermore, this approach was *selected because of the representativeness and uniqueness of their experiences, not because of the generalizability of the findings* [23].

Potential participants were sent an information letter and consent form, giving full details of the study. Once a contacted individual had agreed to participate, an interview was scheduled with each person individually. In total, 19 individuals were approached, and 10 agreed to participate in a semistructured interview. All data were analyzed and reported anonymously. In order to ensure that the participants' current positions were not jeopardized because of the opinions they offered, the names and specific positions of participants were kept confidential.

Data Collection Procedure

Selecting Relevant Sources

We considered documentary sources relevant if they described any information about patient portals in Alberta. They included project management documents, scope or function documents, decision requests, presentations, organizational charts, user manuals, privacy documentation, meeting minutes and follow-ups, contracts, briefing notes, and correspondence. We also included sources that discussed electronic health records (EHRs) or EMRs systems and potential vendors. Sources that were not related to patient portals and/or EHRs or EMRs were excluded. A total of 423 sources were included.

Documents underwent external and internal criticism by MA and were reviewed by PP (co-researcher). External criticism considers the validity of the documents by confirming where the document came from and who had documented the fact that the source existed. Internal criticism looks within the data itself to try to determine truth, even considering the motives of the person providing the data [21]. This process of trustworthiness and credibility was performed for each retrieved source [22,23].

Interviews

The interviews were conducted face-to-face or by telephone by MA and were recorded and transcribed *verbatim*. Participants were given an alphanumeric label to protect their identity. The interviews provided an opportunity to clarify various written sources and eyewitness accounts. A generic description of the positions of the key informants is presented in Table 1.

Table 1. Organizational designation of key interview participants.

Organization and designation of key interview participants	
Alberta Health	• Executive-level participant 1 • Executive-level participant 2 • Executive-level participant 3 • Senior-level participant 4
Alberta Health Services	• IT senior-level participant 5 • Senior-level participant 6 • Executive-level participant 7 • Senior-level participant 8 • Senior-level participant 9 • Clinician senior-level participant 10

Each interview lasted for 45 min and 1 hour. The interview sessions began with clarification of the objective of the study, and a description of the information was sought. The interview questions and discussions focused on the need for a patient portal in Alberta, intended outcomes, facilitators, and barriers to the design and implementation processes, stakeholders, policies and legal factors, and future recommendations.

Data Extraction and Synthesis

The sources were numbered from 1 to 423 for ease of data extraction and referencing. We developed a data extraction form to record any information about the development and implementation processes of any patient portal in Alberta. The final form included items related to source type, date of creation, author position and affiliation, summary of source information, impact of source, possible quotes, and related sources. The primary and secondary sources were grouped by year (2005-2019) and type (planning, decision requests, agendas, minutes, presentations, contracts, scoping documents, and optimization documents). MA conducted the data extraction overseen by PP.

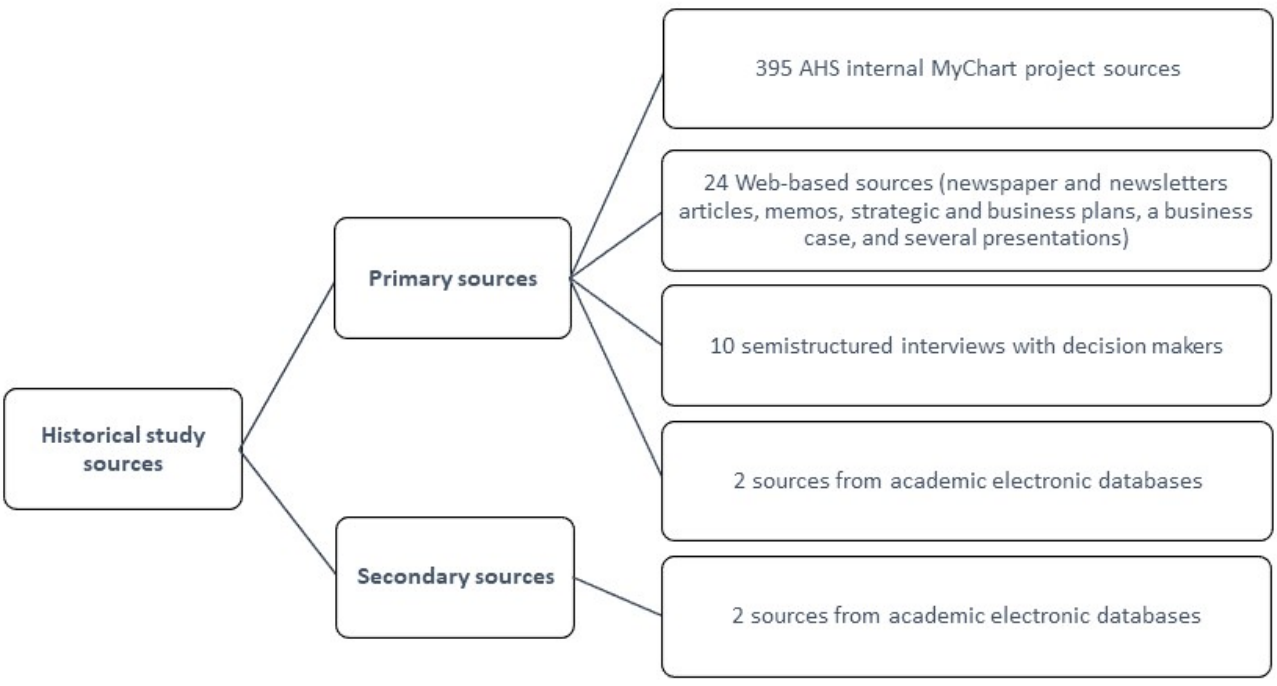
The transcription of the interviews was done by MA and overseen by PP and KO (co-researcher). In historical research, interviews are not analyzed to develop themes but to juxtapose the insights provided by the oral interviews with the *official* documents. The interviews were entered into NVivo (version 11), which was used to track the analytical process through memos and notes. Interview data extraction related to events and actions was connected with patient portals. In addition, potential quotes were identified as being related to various events.

Results

Data Collection Overview

Figure 2 shows the number of selected primary (including interviews) and secondary sources.

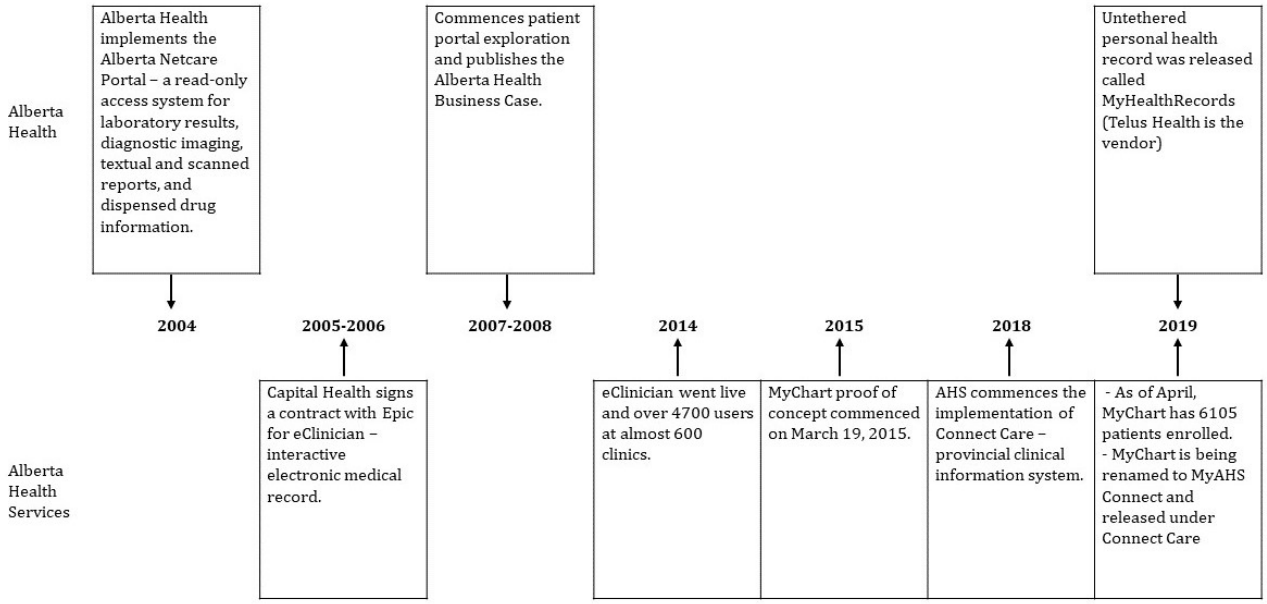
Figure 2. Overview of data collection strategy.



The following sections describe the events based on the extracted data from these sources.

An overview of the Alberta Health and AHS timelines is shown in Figure 3.

Figure 3. Timeline of the development of patient portals in Alberta. EMR: electronic medical record.



The Unmet Need

Before 2005

Alberta's first EMR implementation planning began in 1997 (before the creation of a single province-wide health services delivery organization) with the directive to develop and deliver a single province-wide EMR [24]. This led to the implementation of the Alberta Netcare Portal (ANP) in 2004, which was a read-only access system for laboratory results, diagnostic imaging, textual and scanned reports, and dispensed drug information [25].

2005/06: Capital Health and the Need for a Different Electronic Medical Records

In 2005, Capital Health (1 of the 7 regional health authorities at that time) issued a request for proposals (RFPs) for a different EMR system to replace the ANP system. Epic Systems Corporation (from now on referred to as Epic) was awarded the contract in September 2006. The new EMR system, which was developed and customized for Alberta, was called eClinician.

2007/08: The Alberta Health Business Case

Although Capital Health and later AHS were working on implementing eClinician, AH commenced a patient portal discovery phase based on reports from the United States about the potential for health ITs to be cost saving (business case: personal health portal [Advice to the Minister]). This led Alberta Health to develop a business case that represented the first official document produced by a government body in Alberta that described the need for patients to have access to their personal health records. In addition, patients were *falling through the cracks*, as they were unaware of their test results or if specialist referrals were being made (AH executive-level participant 4).

Although in 2008, AH planned to have a portal deployed within a few years, this did not occur until 2019. Several executive leaders confirmed a description of events during this time when delays resulted from inappropriate planning, changes in vendors, and changes in AH leadership. In addition, individuals involved in managing the personal health record (PHR) project were not equipped with the appropriate technical knowledge as described by an interview participant:

> *So what happened was we ended up in the development mode we're not in the software business. We are not in development business.* [AH executive-level participant 1]

The portal planning undertaken by AHS during 2008 was also seen as a reason for delays:

> *I also think that the culture of AHS has worked against us because they didn't want something separate from their clinical information system. They wanted it to be something that they controlled and part of the software that they would be purchasing. I believe that that culture and that resistance was evident through the whole journey.* [AH executive-level participant 1]

It appears from the key informant interviews and documents that between 2005 and 2008, numerous activities established foundational components for a patient portal in Alberta. It is difficult to judge whether the commitment to meet the identified technology need was rushed or whether the task was more complicated than expected. It took more than 11 years for AH to finally release a PHR system in March 2019 [26]. During these years, an apparent or actual lack of coordination between AH and AHS resulted in tension due to their two patient portal systems appearing duplicative, being established within the same health care system.

2014-2019 Alberta Health Services and the Race to Deliver a Patient Portal

The delays that occurred between 2008 and 2014 were caused by the restructuring that occurred in Alberta when 12 separate health regions and 3 health boards were merged into Canada's first province-wide, fully integrated health system known as AHS. The AHS' patient portal journey was a continuation of what had been initiated by Capital Health (one of the former health regions) by its implementation of eClinician. As AHS was planning the implementation of a provincial inpatient clinical information system, eClinician was meant to serve as a bridge that would ease the gap in terms of the identified need for a provincial EMR and an interactive way for patients and health care providers to access information. In 2014, the eClinician Working Group, with input from various stakeholders, developed a document outlining the AHS health information–sharing prioritization principles that described the need for cost-saving measures based on patients having access to their health care information. In 2014, eClinician went live with over 4700 users at almost 600 clinics or sites [27]. It is important to point out that the AH ANP EMR was still being used in Alberta during the implementation of eClinician.

The eClinician system was intended to support *one person, one record*, and had the capacity to deliver the MyChart patient portal through which patients could access their record, communicate with their provider, and book and cancel appointments. The MyChart PoC, led by a clinical Working Group (MyChart Working Group) reporting to the AHS Ambulatory Oversight Council, was planned in phases that were meant to reflect a forward-thinking plan to align with both the AH PHR and the upcoming AHS Clinical Information System.

> *As a small project team what we did is we created a Clinical Guidance Working Group it was called the MyChart Working Group and on there we had a number of different users on board. We had physicians, nurses, allied health, from different sites and we had managers on there as well.* [AHS senior-level participant 8]

The MyChart PoC implementation was planned as an incremental change in order to minimize resistance by clinicians and patients (eClinician prioritization principles). It was decided that at the conclusion of the PoC, the Working Group would validate the solution against physician's expectations, assess the true performance of the patient portal in a controlled environment, identify areas for improvement, identify implementation tips and traps, and lessons learned, measure key performance indicators and determine the return on investment (eClinician link PoC scope August 21, 2015).

The results from the PoC would be used to determine any future approaches for deployment, long-term expansion, or modifications of the health IT. Thus, between 2014 and 2019, AHS planned, developed, customized, and implemented a patient portal—a process that has been characterized by intricate agenda-setting and decision-making processes. The following sections will describe the technology and what needed to be done to commence the PoC and the subsequent successful implementation and adoption.

MyChart: The Change

The AHS MyChart was a customizable Web app that offered patients easy access to their medical records via controlled access to the same eClinician medical records used by their physicians. It provides self-service functions that have the capacity to reduce administrative costs and increase patient satisfaction. It also offers various features that organizations can select to meet their identified requirements (MyChart Recommendations: Core Features). Each function is implemented based on the need and cost.

Patients participating in the MyChart PoC were able to view their health summaries (problem list, medications, allergies, and immunizations), laboratory and diagnostic imaging test results, previous and upcoming appointments, and letters sent from clinics they have attended. In addition, patients were able to send nonurgent messages to their health care team, request appointment dates and times, complete health assessment questionnaires, and enter information (ie, vital signs or blood glucose measurements).

Before deciding on how the PoC will evolve and if MyChart was the right technology, the work was precipitated by extensive engagement with various stakeholders, as described in Figure 4 (MyChart summary report). This wide consultation process allowed the MyChart Working Group to build credibility in the work they undertook.

Before the PoC started, the MyChart Working Group had to plan and consider all possible variables, tasks, and situations that might arise, and this work was guided by the Epic team (weekly status report for eClinician foundation—eClinician MyChart). The planned PoC phases are shown in Figure 5.

The PoC was deployed by ensuring that the best *bang for our buck* was achieved by only working with health care providers that were willing to be part of the study and implementing functions that were going to be utilized (MyChart summary report). The vision for MyChart PoC was *better health, powered by information, supported by technology* (MyChart monthly CIS project status report, August 2015). The estimated PoC implementation cost (more than US $873,600) was based on the cost of staff to support the PoC, and the purchase of the required software and hardware (MyChart PoC scope, September 9, 2015). By April 22, 2015, the PoC was up and running. The initial goal was to engage 500 patients participating across clinics, with significant measurement and benefits realization activities.

Figure 4. Stakeholder involvement in the MyChart planning and implementation. AHS: Alberta Health Services; IT: information technology.

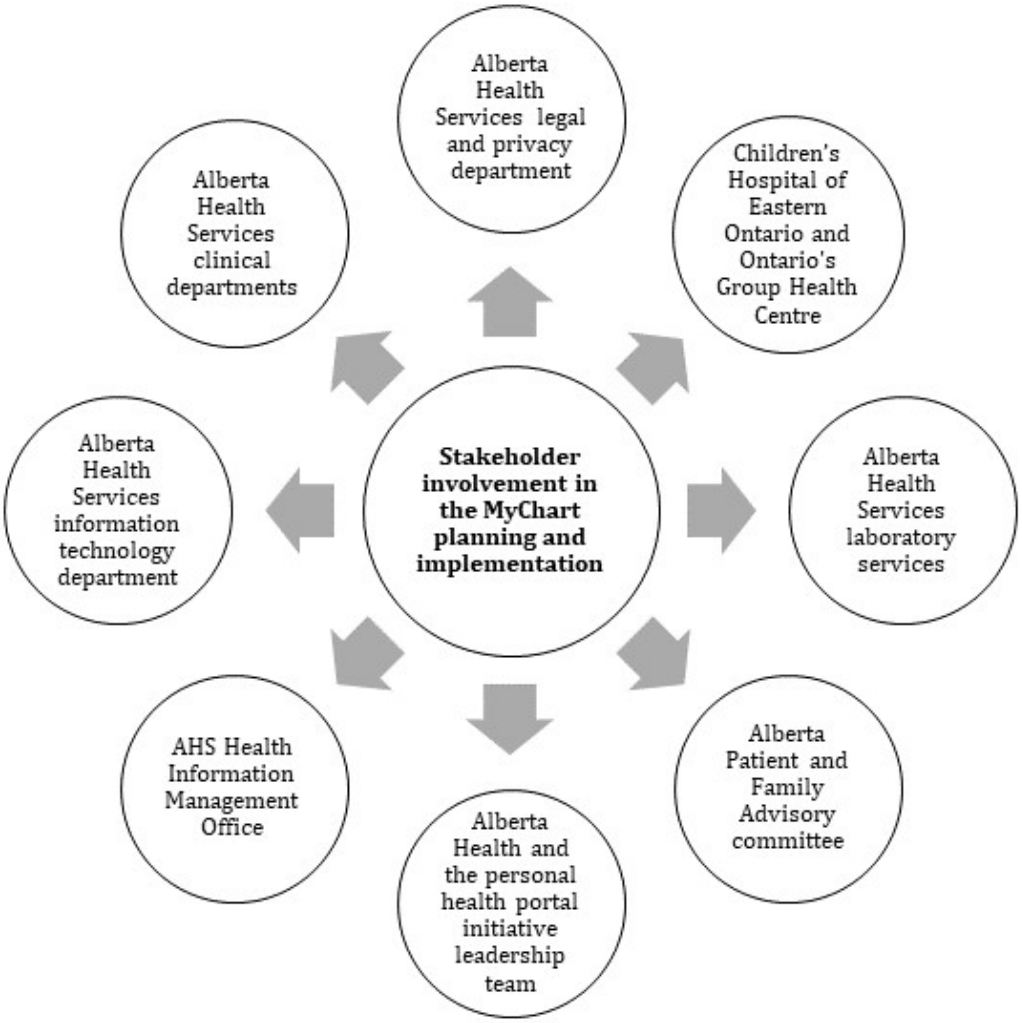
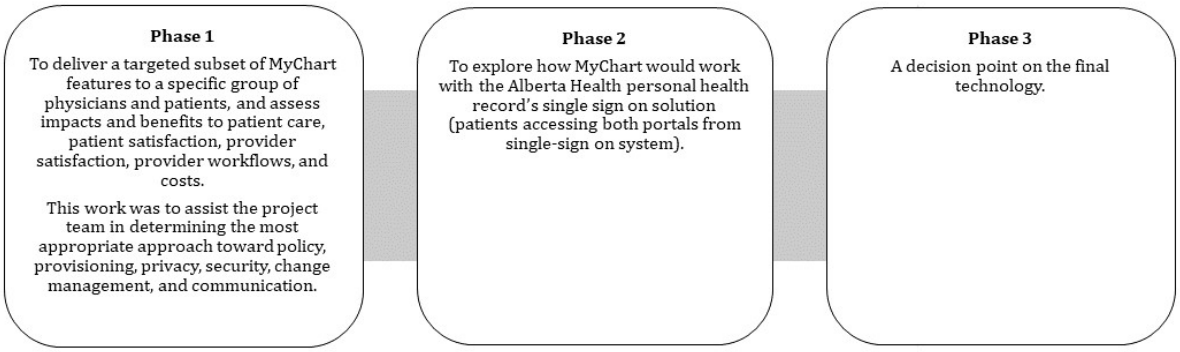


Figure 5. Planned MyChart proof of concept phases. PHR: personal health record.



Making the Case for MyChart

The launch of the MyChart PoC required several components that were determined to be crucial for success, such as the recruitment of clinics and health care providers, customization of the portal, and an understanding of how MyChart would align with the AH PHR initiative.

The MyChart Working Group established project management principles to guide them through the planning and implementation process, guided by charter principles as shown in Textbox 1 (eClinician: Lessons learned in context MyChart project).

Textbox 1. MyChart proof of concept charter principles.

- Keep it simple
- Keep the focus on the patient
- Have clear communication with each other and with the task force group
- Let us work together, collaborate, and take joint responsibility for getting it done (no finger pointing)
- Listen to Epic! (Let us not waste time reinventing the wheel, Alberta Health Services is not all that different)

In making the case for MyChart, the MyChart Working Group outlined several cost-saving assumptions, such as a decrease in no-show appointments due to MyChart reminders and the opportunity to implement Web-based scheduling. In addition, with secure messaging, there would be less need for mailing letters to patients, and fewer visits to the emergency department for nonemergent issues, as patients would have immediate access to their health care team. Given that patients would have access to their medical records and test results, the number of follow-up visits was also expected to decrease (eClinician MyChart PoC).

Costs were calculated for the license fees and maintenance of the MyChart app. For each new user who accesses MyChart at least once, there would be an ongoing cost of US $4.05 per year (eClinician MyChart PoC). These costs were offset by the savings realized through the avoidance of unnecessary medical visits.

The Working Group members considered the MyChart PoC as a *bottom up initiative* based on the notion that clinicians had asked for their patients to have access to their medical records (MyChart Epic care AMH, December 3, 2015). Furthermore, the process of *customizing* the patient portal from the American version to a version that worked within the legislation and privacy requirements in Alberta was a long process. There were many decision requests to change how the processes and functions worked, with the intent to create a system that was patient centered and would meet all legislative requirements of Alberta's Health Information Act.

Gaining Support from the Providers

As described by AHS senior-level participant 9,

> *because we wanted to be successful, we sort of chose a group of willing clinics and physicians who we knew were knowledgeable, that reassured us that the clinic was committed and that they were willing to help us. The Working Group didn't want to force anybody that did not want it to participate in the PoC.* [AHS senior-level participant 9]

The recruitment of participating health care providers was a challenging process due to the uncertainty of what a patient portal would mean for them:

> *There were some physicians who were absolutely horrified at the idea of patients being able to view their own results and their own information especially in the case where some decisions were made to release lab results immediately. They were horrified at the idea that the patient might see the results before the physician had a chance to see the result.* [AHS senior-level participant 9]

Health care providers were concerned because MyChart would not replace anything but instead add to the workflow. The MyChart Working Group wanted to ensure that the initiative would be seen as useful and necessary for the enhancement of patient-centered care. The Working Group decided that the best way to commence the PoC was to find health care providers who were given the autonomy to decide which patients were appropriate candidates for MyChart.

During the promotional and exploratory processes for participating clinics, the Working Group selected 5 clinics in Edmonton to participate in the PoC, as shown in Textbox 2, and offered the opportunity to the health care providers from these clinics to participate or not to participate in the PoC. As of April 2019, MyChart was used at 10 clinics.

Textbox 2. MyChart proof of concept participating clinics.

- Lynnwood Family Medicine Clinic
- Alberta Health Services (AHS) Inflammatory Bowel Disease (IBD) Clinic
- AHS Kaye Edmonton Clinic (KEC) Multiple Sclerosis
- AHS KEC Diabetes
- AHS KEC Rheumatology

Navigating Between the Two Patient Portals

The leading concern about the future of MyChart and the need for a PoC hinged on processes related to deciding how the AH and the AHS patient portals were ultimately going to *work together*:

> *The AH PHR is essentially access to read-only database of health information. If you want to break it down to most basic stuff, PHR is for reviewing health information, while MyChart is a tool for viewing health information. These are all things that the PHR can never do. It is helpful to say that they*

> *are two different things, the problem is that if the patients are coming into MyChart and they can do all these things, why would they bother going to the PHR? That's the problem.* [AHS IT senior-level participant 5]

As the introduction of the MyChart PoC idea, tensions between AH and AHS in terms of the portal delivery have increased, leading to a *competition-type* delivery of service, as the two systems appeared duplicative.

> *There were a lot of pull and retain, and a lot of political sensitivities given the amount of money that Alberta Health, the time and effort and money had invested in the PHR. So to get the MyChart® pilot going, we had to escalate it up out of our zone to the IT leadership and medical leadership at AHS first. Then, we had to get permission from the deputy minister level of Alberta Health to do our pilot. We weren't allowed to advertise.* [AHS executive-level participant 7]

From the early planning days of the MyChart PoC, it was determined that the MyChart portal provided more functionalities than the planned PHR.

> *We knew that we could offer more complete functionality using the Epic MyChart, basically out of the box. So the public portal that they're building, there was a lot of configuration, they had to build the software. There was a lot of configuration to that kind of thing. And then the software platform that they were building on it was no longer being supported by Microsoft. So we knew that we could surpass their functionality out of the gates in some ways using eClinician MyChart.* [AHS executive-level participant 7]

Regardless of what was going to be decided, there was a feeling that AH has had the opportunity since 2009 to deliver a patient portal. The sentiment expressed by some participants was that money has been spent, and patients were yet to have access to their health care records. MyChart was seen as a portal that had the potential to be delivered on the identified need.

> *And for clinicians and patients alike this is a real game changer. There's something that really changes the way we operate in health care. It takes responsibility for health care and it gives it to me, to the patients, they can take responsibility for their own health care, which in theory is everything driven by the physician and dictated by the physician in terms of why he thinks the patients need to know and understand.* [AHS senior-level participant 9]

The planning for MyChart had to incorporate components as to how the two portals will eventually work together.

Proof of Concept Scope and Functions

The most challenging component of the implementation process was determining the MyChart functions, as there were many options but limited funding and personnel. Organizational policies had to be established for a help desk, customer service requests, key performance indicator report analyses, a MyChart utilization dashboard, and a *go-live* strategy (Epic MyChart project plan, February 2, 2015). There were many functionalities from which to select; therefore, the Working Group had to decide what was important during the PoC.

Then, 1 participant described:

> *we just sat down in committee meetings and went through the potential futures and whether it was feasible and useful and did we want. And that's how we kind of made our decisions.* [AHS executive-level participant 7]

The Working Group focused on developing the sign-up process and implementing the following functions: two-way communication, real-time scheduling, release of test results, proxy access, notifications, appointment scheduling, and questionnaires, as described in Table 2.

Decisions on how to proceed were made collaboratively as the Working Group saw this as an opportunity to bring a patient portal in Alberta. As 1 participant described:

> *so it is this step wise process that we went through and we just sat down in committee meetings and crack through the potential futures and whether it was feasible and useful and did we wanted and our environment. And that's how we kind made our decisions* [AHS executive-level participant 7]

Therefore, some functions were selected, whereas others were not. Furthermore, when the MyChart PoC was planned, it was decided that the goal would be to offer MyChart to about 500 patients. However, as the PoC evolved and more health care providers decided to participate, it became clear that the milestone of 500 patients was no longer adequate (eClinician MyChart PoC). After considering the positive feedback to date, the AHS agreed to allow for the number of patients participating in the MyChart PoC to be increased up to 5000.

Although there were a great number of anticipated challenges with bringing a patient portal in Alberta, AHS and the MyChart Working Group concluded that not proceeding with this PoC was not an option.

Table 2. Planning and development of functions during the MyChart proof of concept.

Function	Planning and development process
Sign up process	• The sign-up process was a source of significant grievances by patients and health care providers, "because we put so much security on it and made them use passwords, we made it quite complex" (AHS clinician senior-level participant 10). • A two-factor authentication process was designed to comply with privacy legislation. Some never tried to create a MyChart account again after an unsuccessful initial attempt. • When MyChart was first introduced, the sign up was based on a process whereby a provider would print a letter containing a MyChart activation code, and then the patient would take it home and sign up for the account at his or her convenience. Although this process seemed simple, many patients had issues with the code, remembering to set up the account, or losing the printed paper. • There were numerous discussions on how to improve this process and make it more efficient (MyChart activation workflows—pros and cons). • After considering the various options for sign up and soliciting advice from Epic, the MyChart Working Group decided on an email process. • Once a provider had introduced and discussed MyChart with the patient, and he or she agreed to sign up, the provider clicked the MyChart status icon on the patient header in eClinician [Decision Request 11: Create option for MyChart activation letter to be sent to patients via email]. This initiated an email containing an activation letter that was directly sent to the patient's email address documented in MyChart. The email included general instructions about MyChart and the patient's MyChart activation code. The patient had to enter his or her personal health care number, date of birth, and activation code. • The sign-up process included attaching a label to all patients offered MyChart, which indicated whether they were *active* (account used on regular or as needed basis), *inactive* (patient signed-up but never used), *pending* (a patient starts the sign-up process but does not complete it), and *declined* (patient declines to sign up). These labels were as a way of keeping track of patients (Decision Request 238: Add MyChart status to patient header).
Two-way communication	• MyChart included two-way patient-provider secure communications for nonemergent issues. This function required many modifications and decisions because it had not been tested before the PoC, and participating clinicians had expressed skepticism. It was also one of the main reasons many health care providers within the participating clinics decided not to participate in the PoC. Providers assumed that they would be inundated with a large number of messages, and there was no compensation plan in place to remunerate them for their time. • This function was seen as time saving not only for providers but also for patients. • To effectively evaluate the two-way communication, clinics had to decide on the message routing, and how responses would be managed, although the clinics agreed to a process, and each adapted to meet their needs. In some clinics, clinicians monitored the messages directly, whereas in other clinics, nurses or the front office staff managed messages.
Real-time appointment scheduling	• For the real-time scheduling function to be applied, participating clinicians had to enter their availability in the system, which would allow patients to select from the available slots. Once a patient selected a slot, the clinic's office would receive a message. If the booking was done incorrectly or the slot changed, the office staff would call the patient and modify the booking. • This function allowed patients to not only book appointments, but also cancel them at their convenience. Although this function was available to all clinics during the PoC, only 1 clinic (community-based family clinic) gave their patients access to real-time scheduling.
Release of test results	• Release of test results required the lengthiest deliberations because it needed to comply with the legislative requirements of the Health Information Act. • Three options were discussed: auto release (immediate release of all nonsensitive results), time-delayed auto release (some results to be released after a 7-day delay), and no release (results of sensitive tests with the potential for security and privacy concerns). • Each was explored, although Epic recommended automatically releasing as many results as possible, the MyChart Working Group decided that results would be released after 10 days. • Patients would receive an email notification once the results were posted in MyChart.

Function	Planning and development process
Proxy access	• Alberta Health Services determined that although proxy access was challenging due to the Health Information Act, this function would be piloted. • Proxy access allowed MyChart users to permit others to access their MyChart record by establishing a proxy relationship (eg, parent to child, adult to elderly parent). Proxy access was seen as *breaking new ground* and therefore required careful monitoring and clear guidelines. "So proxy was essentially a precedent issue," as no service provided by Alberta Health Services had this available electronically (AHS IT senior-level participant 5). • When initially introduced during the PoC, proxy access was granted through a manual paper-based process that was found to be very cumbersome and time-consuming. It was later replaced by an electronic proxy access via eClinician. Another issue was that in order to receive proxy access to another patient's MyChart account, the designated proxy had to first have their own personal MyChart account. This was not always possible, as many family members requesting proxy access were not part of the same clinic that was on eClinician. • The Working Group established a *shell* MyChart account containing no personal health information for those that have been approved proxy access to another patient's MyChart account, but do not have their own MyChart account (AHS IT senior-level participant 5). Although clinics found the proxy sign-up process cumbersome, the benefits of having proxy access appeared to outweigh these concerns, as patients reported positive benefits.
Questionnaires	• Although several participating clinics wanted to utilize MyChart for questionnaires and surveys, the MyChart Working Group decided for only 1 specialty clinic to proceed with this function. • It was agreed that if this function was deemed successful, it would be recommended for wider implementation.

Discussion

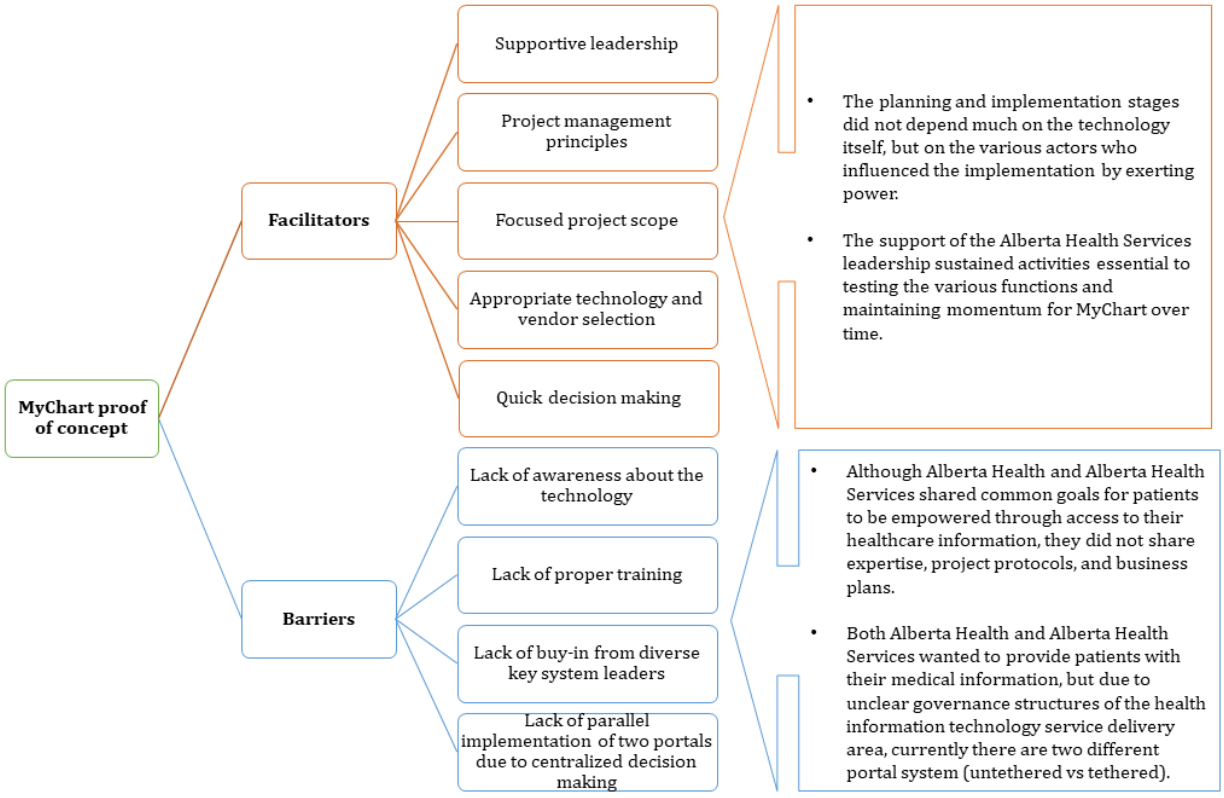
Principal Findings

We have presented an overview of the various environmental and contextual factors that influenced the MyChart PoC during the planning and implementation stages. The MyChart PoC was a complex undertaking of design transformation, adjustment, implementation and deployment, collaboration and compromise, and problem solving with the intent to learn about the necessary elements needed to ensure successful wider implementation. The PoC leadership team agreed that the "technology was the easiest part," as they focused on foundational features with very minimal customization (AHS IT senior level participant 5). The problem-solving processes were impacted by various environmental, social, and professional factors.

Despite extensive research in the area of patient portals, implementation and subsequent adoption of these systems is a convoluted process, which tends to impact settings and practices in different ways [28]. This complexity is clearly shown in MyChart PoC. Furthermore, regardless of the type of setting, these systems require a substantial investment of time, resources, and determination, and not all care settings have the same capacity to contribute equally. As the PoC proceeded, the Working Group saw a difference in the enrollment numbers between the clinics that alluded to differences in uptake.

Through this research study, we wanted to determine what it took to implement a patient portal and what were some of the elements that are necessary not only for successful implementation but also for the successful adoption and optimization of the technology. The review of primary and secondary sources and insights from the interviews explicated several facilitators and barriers that affected the process, as described in Figure 6.

Figure 6. Facilitators and barriers that impacted the MyChart proof of concept during the planning and implementation stages. AH: Alberta Health; AHS: Alberta Health Services.



MyChart Proof of Concept Facilitators

Supportive leadership and quick decision making, project management and focused project scope, and appropriate technology and vendor selection were some of the facilitators that allowed for the commencement and growth of the PoC.

Supportive Leadership and Quick Decision Making

The MyChart Working Group comprised leaders who had the capacity to make decisions in support of making the implementation more manageable and without unnecessary delays. It was not only that the leadership was effective, but they also used real and relevant cases to make the case for the portal and garner support. These observations were also confirmed by an interview participant:

> *I found that in some way, so from concept to deployment it took us what around 6 to 8 months with MyChart. And the reasons I think we were successful is we had strong leadership at the level of the MyChart pilot. And there was medical leadership but also operational leadership. We explicitly had use cases. So we developed a little framework about what are you actually going to use it for, explicitly. That is very helpful and having understanding of the high level concepts. And then we were prepared to make quick and rapid decisions and live with those decisions. And we had clinical input to help us make the decisions.* [AHS executive-level participant 7]

Leadership was able to try improvements and become leaders in innovations, as they had

> *own internal executive approval to kind of go for it and try new things and create some precedent.* [AHS IT senior-level participant 5]

Furthermore, the working Group

> *had to set the stage when it came for the precedent setting stuff since they considered this project as a PoC and they were given a little bit more leeway.* [AHS IT senior-level participant 5]
>
> *The project leadership was able to create a pull for the technology and created an environment where the demand was greater than the supply.* [AHS IT senior-level participant 5]

The findings show that the planning and implementation stages did not depend much on the technology itself, but on the various actors who influenced the implementation by exerting power.

> *The proof of concept was ultimately a learning exercise.* [AHS IT senior-level participant 5]

As described by AHS executive level participant 7:

> *We were successful because we had strong leadership at the level of the MyChart pilot.*

The support of the AHS leadership sustained activities essential to testing the various functions and maintaining momentum for MyChart over time (eClinician: Lessons learned in context MyChart project).

Project Management and Focused Project Scope

The MyChart Working Group developed the PoC based on project management principles. These principles were upheld

through various meetings and discussions. Furthermore, the Working Group developed many discussion documents to identify the most suitable scope for the MyChart PoC. As confirmed by AHS IT senior-level participant 5:

> *first got scope, and once we got clear understanding of who is going to be involved, resources. So scope, resources, finance, and then we could build the schedule. That schedule looked at basically at half dozen clinics and we rolled out everything in a very short period of time.*

Appropriate Technology and Vendor Selection

The MyChart Working Group saw Epic as the necessary guide in this PoC as they possessed the skills and knowledge on how to implement this type of technology. The assumption was that Epic had done this in many different settings and they had the expertise to understand what was required and what AHS needed to do in order to have a successful PoC. The relationship with Epic was not only for the implementation of eClinician, but also for future partnerships.

> *So it wasn't really because of the former relationship with them on eCLINICIAN deployment on anything like that, it was more of a perceived ability for them to help us in the future as a partner to deploy the full Connect Care suite of options.* [AHS executive-level participant 7]

The partnership between the MyChart Working Group and Epic has continued and grown.

MyChart Proof of Concept Barriers

The main barriers were lack of awareness about the technology, proper training, buy-in from diverse key system leaders, and parallel implementation of two portals due to centralized government decision making.

Lack of Awareness About Technology

Although the project team decided not to impose the MyChart PoC on any clinic, various documents show that there was a substantial "lack of awareness" (AHS senior-level participant 6) about what patient portals are and what they would do.

There were barriers relating to communicating with the clinics and getting buy in. There was the fear that getting patients their own information would lead to increased work for the physicians in particular. [AHS executive-level participant 7]

Therefore, there was a fear

> *that patients wouldn't be ready, would get really scared and this would cause a lot of extra work for physicians.* [AHS executive-level participant 7]

Lack of Proper Training

The MyChart Working Group acknowledged that the clinic staff should have received more detailed training. Clinicians perceived MyChart as an add-on and thus did not see training as something needed. This issue was amplified by the fact that the Working Group, due to the small funding, did not have the capacity to establish dedicated staff to assist with any technical and user issues. Although training was not deemed as valuable, as the PoC progressed, clinicians and other clinic staff realized the need for proper training on how the portal works and how patients interact with the system.

Buy-In From Diverse Key System Leaders

Buy-in was a challenge due to

> *fear that giving patients their own information would lead to increased work for the physicians in particular.* [AHS executive-level participant 7]

Furthermore, there was a lack of different key system leaders and champions. The MyChart Working Group comprised influencers, but they needed additional leaders to expand their spheres of influence.

> *I think you need for any successful innovation you need champions both within and outside the system but within the system is absolutely critical and fundamental. If you don't have internal champions, it's not going to go anywhere. And I think you really need in most instances health clinician champions.* [AH executive-level participant 2]

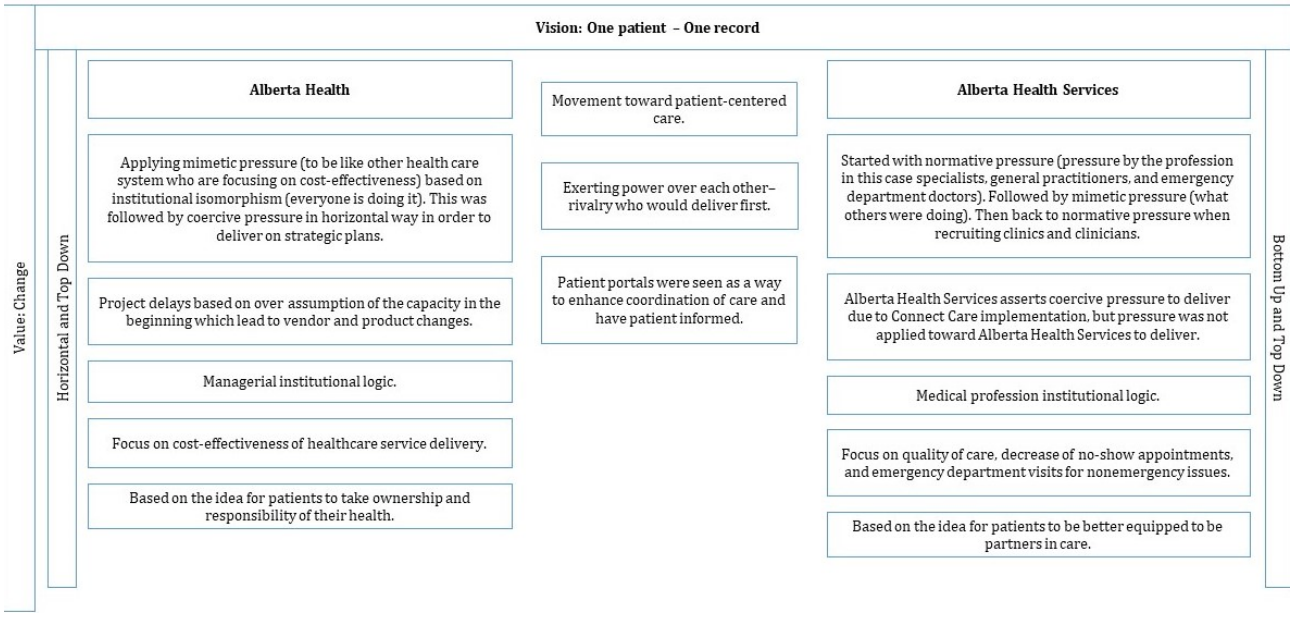
Leadership change also affected the MyChart PoC in negative ways. There were several documents called *handover succession checklists*, which described the complex process of anytime individuals changed (Enhancement PoC eClinician portals: MyChart and eClinician link—Handover succession checklist).

> *So when you have senior leader change in charge of a project that makes for a challenge as well in terms of transition thing.* [AHS clinician senior-level participant 10]

Parallel Implementation of the Two Patient Portals

Although AH and AHS shared common goals for patients to be empowered through access to their health care information, they did not share expertise, project protocols, and business plans. In addition,, the two approaches described in Figure 7 show the impact of the parallel implementation of the two portals.

Figure 7. The vision to change. AHS: Alberta Health Services.



As described by an interview participant:

> *there are two because neither have fully developed themselves out to be the one that is preferential and two, not everyone was involved.* [AHS senior-level participant 6]

Convergence of the two portals seems to be a natural progression, described as follows:

> *Convergence has to happen. Someone somewhere eventually has to figure it out.* [AHS executive-level participant 7]

Both AH and AHS wanted to provide patients with their medical information, but due to unclear governance structures of the health IT service delivery area, currently there are two different portal systems (untethered vs tethered) [29]. These bureaucratic governance complexities will continue to impact how services are planned and delivered in Alberta until there is a better governance structure [30].

The MyChart PoC facilitators and barriers show the complexities that exist within the decisions to implement health IT within different health care settings.

Comparison With Current Literature

Several studies have focused on identifying facilitators and barriers to patient portal implementation, but only a few have touched on understanding the organizational impact [31-33]. Kooij et al [31] focused on understanding how organizational factors impact the implementation process of patient portals by focusing on several hospital settings and identified a number of facilitators and barriers. Their findings are similar to what we found in terms of a lack of perceived value and willingness to change by health care providers. These findings are also similar to those of Koivunen et al [32] and McGinn et al [33]. However, these latter studies focus on understanding the facilitators and barriers from the perspectives of different stakeholders (providers, managers, IT providers, and patients), whereas our focus has been on understanding the internal organizational facilitators and barriers that impact the planning and implementation stages of a patient portal system. Our intent was to understand the internal drivers of change and how decisions were made to support implementation.

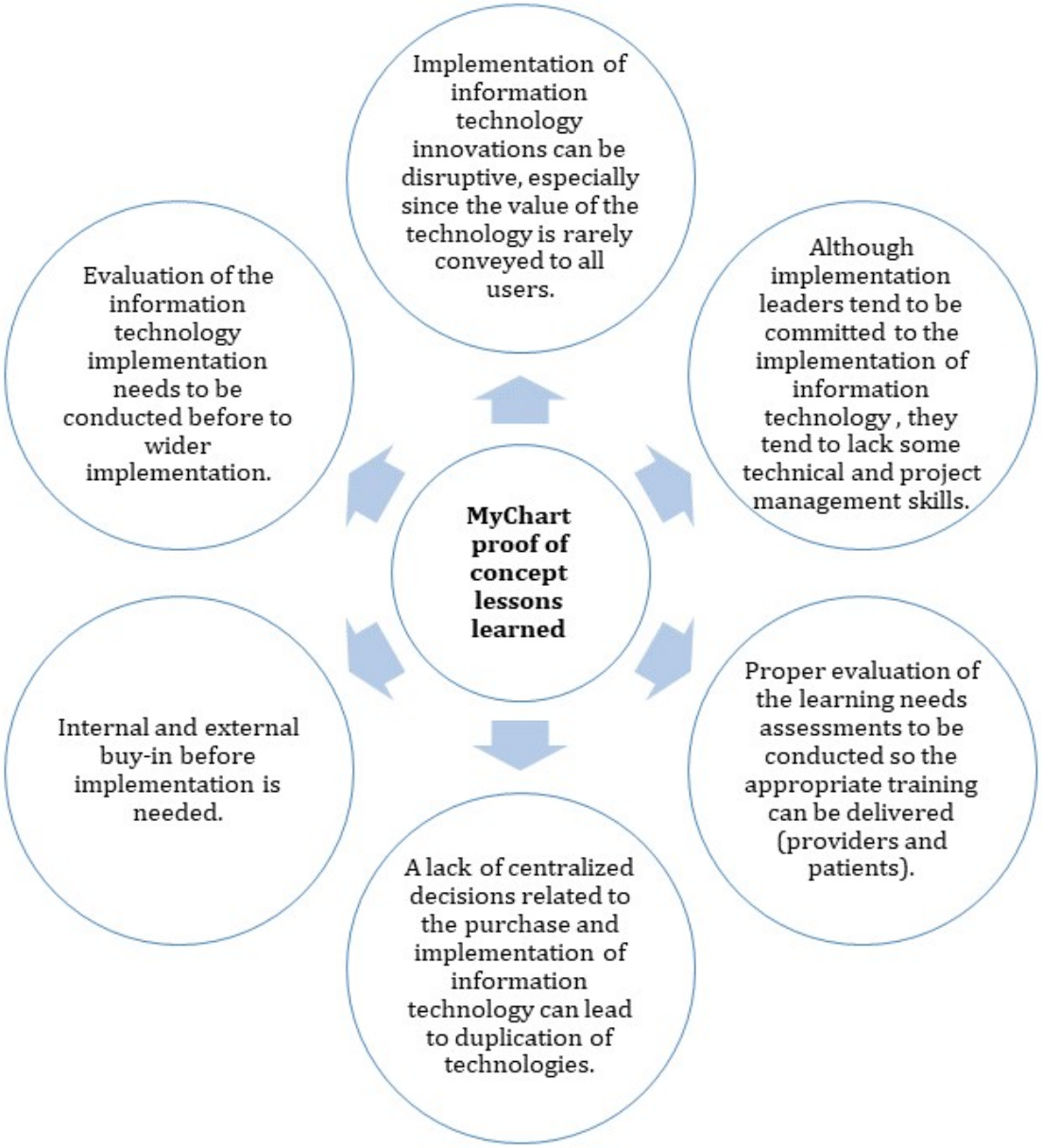
In this study, we found that incorporating project management principles led to a more focused scope that was aligned with the limited funding. These principles (vision and mission, objectives, standards of engagement, intervention and execution, organizational alignment, and measurement and accountability) allowed the implementation team to maintain a detailed record and track progress in real time. The Working Group incorporated project management principles for each project stage (planning, development, implementation, optimization, evaluation, and adoption). Studies by Richer et al [34] and Aubry et al [35] confirm that projects that have the potential to impact organizational change provide benefit not only by incorporating project management principles but also by establishing a project management office and central decision making to improve resource allocation.

According to the US Office of the National Coordinator for Health Information Technology, *the challenge of narrowing a large field of available options to a manageable number of vendors can be daunting, but it is a critical step* [36]. It is not only about selecting the best vendor but also selecting a vendor that *is willing to make commitments to gain new business* [37]. In this study, we were able to gain an in-depth understanding of how the vendor selection process evolved. The vendor section was based on the perceived ability of Epic to help in the future as a partner to deploy the provincial clinical information system.

Implications and Lessons Learned

Although many aspects of what occurred during the MyChart PoC processes are considered common with the implementation of health ITs, there are several features with broad implications for planning and delivering patient portals in a large public health care system that were evident during the MyChart PoC in Alberta, as described in Figure 8.

Figure 8. MyChart proof of concept lessons learned. IT: information technology.



First, the implementation of IT innovations can be disruptive, especially as the value of the technology is rarely conveyed to all users [38,39]. *Value, by definition, includes both costs and benefits* [40]. This is confirmed by several reviews [40-43]. One review concluded, the *human element* is critical to health IT implementation [42]. Although these IT innovations are seen as a *key component of health care transformation to reduce costs and improve quality*, their adoption *proceeds at a snail's pace* [40]. Furthermore, there tends to be a lack of properly *selecting which IT projects are of most value* [43]. Selecting the right technology requires an understanding of the value that the technology has the capacity to deliver and the ability to convey that value to all end users [44,45]. As the literature points out, health IT implementation should be considered as a complex intervention and that *complex interventions may work best if tailored to local circumstances rather than being completely standardized* [46]. Second, although implementation leaders tend to be committed to the implementation of IT, they tend to lack some technical and project management skills [47]. These types of implementation require appropriate skills to ensure successful utilization [46,48]. Third, proper evaluation of the learning needs assessments to be conducted so that appropriate training can be delivered (providers and patients) [49-52]. Quick presentations or Web-based technical training did not seem to resonate with the health care providers during the MyChart PoC. Fourth, a lack of centralized decisions related to the purchase and implementation of IT can lead to duplication of technologies [47,53-55]. Fifth, internal and external buy-in before implementation is needed [47,56,57]. Studies have shown that without the proper buy-in from providers, patients lack information about the opportunities to access their medical records [58]. Without the proper buy-in, providers and decision makers tend to doubt the usefulness of the technology [39]. Sixth, evaluation of IT implementation needs to be conducted before wider implementation [59,60]. Furthermore, evidence from pilots is rarely used when planning the implementation

on a wide scale [60]. According to Houston et al [61], *the domain of Health Informatics is at risk for too rapid implementation as external pressures continue to promote adoption*. Although decisions are made to invest in the purchase and implementation of patient portals, the evidence of utilization and adoption has produced mixed results [41,62,63]. There is a rapid desire to implement patient portals without clear evidence that these technologies have the desired impact on the target populations in terms of effectiveness and safety [61].

The findings in this study highlight the importance of understanding internal organizational and decision-making approaches that have the capacity to hinder the planning and implementation of patient portals. It shows how organizations decide on IT investments, the intricacies of the decision processes, and factors affecting decisions at each stage to provide better future preparations for the successful implementation of technologies. Furthermore, our findings also document the effect of various social and political spheres on the development and implementation of MyChart and identify key factors that government and health care organizations may wish to consider before funding IT in health care.

Limitations

There are several limitations to this study. First, we were unable to receive archival information from AH to describe their patient portal development process in its entirety. Second, although many efforts were made, we were unable to interview any AHS individuals who were considered IT technical experts.

Conclusions

Implementing patient portals is a complex undertaking, as “it’s much more about the people who are using it that actually can make an impact on care” (AHS senior-level participant 6). The results of this study document the effect of various social and political spheres of influence on the development and implementation of MyChart and identify key factors that government and health care organizations may wish to consider before funding IT in health care. This study supports decision makers in understanding and managing the necessary organizational change and managing the individual expectations when implementation technologies have different types of usage by different groups.

Acknowledgments

The authors would like to thank each of the interview participants who took the time to participate in this study. Their participation ensured that we were able to verify the validity of various documents. They would also like to thank the MyChart Working Group for sharing the documents with us. Without access to these internal documents, the story could not be told.

Conflicts of Interest

None declared.

References

1. Sheikh A, Sood HS, Bates DW. Leveraging health information technology to achieve the 'triple aim' of healthcare reform. J Am Med Inform Assoc 2015 Jul;22(4):849-856 [FREE Full text] [doi: 10.1093/jamia/ocv022] [Medline: 25882032]
2. Bates DW. The quality case for information technology in healthcare. BMC Med Inform Decis Mak 2002 Oct 23;2:7 [FREE Full text] [doi: 10.1186/1472-6947-2-7] [Medline: 12396233]
3. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med 2006 May 16;144(10):742-752. [doi: 10.7326/0003-4819-144-10-200605160-00125] [Medline: 16702590]
4. Black AD, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. PLoS Med 2011 Jan 18;8(1):e1000387 [FREE Full text] [doi: 10.1371/journal.pmed.1000387] [Medline: 21267058]
5. Alotaibi YK, Federico F. The impact of health information technology on patient safety. Saudi Med J 2017 Dec;38(12):1173-1180 [FREE Full text] [doi: 10.15537/smj.2017.12.20631] [Medline: 29209664]
6. Kaplan B, Harris-Salamone KD. Health IT success and failure: recommendations from literature and an AMIA workshop. J Am Med Inform Assoc 2009;16(3):291-299 [FREE Full text] [doi: 10.1197/jamia.M2997] [Medline: 19261935]
7. Xue Y, Liang H, Boulton WR. Information technology governance in information technology investment decision processes: The impact of investment characteristics, external environment, and internal context. Manag Inf Syst Q 2008;32(1):67-96. [doi: 10.2307/25148829]
8. Bryant A, Black A, Land F, Porra J. Information systems history: what is history? What is is history? What is history? … and why even bother with history? J Info Tech 2013 Mar;28(1):1-17. [doi: 10.1057/jit.2013.3]
9. Mason RO, McKenney JL, Copeland DG. An historical method for MIS research: Steps and assumptions. Manag Inf Syst Q 1997 Sep;21(3):307-320. [doi: 10.2307/249499]
10. Christy TE. The methodology of historical research: a brief introduction. Nurs Res 1975;24(3):189-192. [Medline: 1039626]
11. Glass L. Historical research. In: Brink PJ, Wood MJ, editors. Advanced Design in Nursing Research. Second Edition. Thousand Oaks, California: Sage Publications; 1998:356-374.

12. Fitzpatrick M. Historical research: the method. In: Munhall PL, Boyd CO, editors. Nursing Research: A Qualitative Perspective. Fifth Edition. Sudbury, MA: Jones & Bartlet Learning; 2012:359-371.
13. Mitev N, de Vaujany F. Seizing the opportunity: towards a historiography of information systems. In: Willcocks LP, Sauer C, Lacity MC, editors. Enacting Research Methods in Information Systems: Volume 1. Hampshire, England: Palgrave Macmillan; 2012:235-266.
14. Government of Alberta. Ministry of Health URL: https://www.alberta.ca/ministry-health.aspx [accessed 2017-02-05]
15. Suchowersky A, Suchowersky O, Duckett SJ. Can Alberta's primary care networks provide any lessons for Medicare Locals? Med J Aust 2012 Jan 16;196(1):27-28. [doi: 10.5694/mja11.11109] [Medline: 22256923]
16. Alberta Medical Association Primary Care Alliance Board. PCN Evolution. Alberta, Canada: Alberta Medical Association; 2013 Dec. PCN Evolution: Vision and Framework. Report to the Minister of Health URL: http://www.pcnevolution.ca/SiteCollectionDocuments/PCNe%20Overview/PCN%20Evolution%20Vision%20and%20Framework.lrg.pdf [accessed 2018-03-12]
17. Alberta Health Services. About AHS URL: https://www.albertahealthservices.ca/about/about.aspx [accessed 2017-02-05]
18. Lewenson S, McAllister A. Learning the historical method: step by step. In: de Chesnay M, editor. Nursing Research Using Historical Methods: Qualitative Designs and Methods in Nursing. New York, NY: Springer Publishing Company; 2015:1-23.
19. Mages KC, Fairman JA. Working with primary sources: an overview. In: Lewenson SB, Herrmann EK, editors. Capturing Nursing History: A Guide to Historical Methods in Research. New York, USA: Springer Publishing; 2008:129-148.
20. Toland J. Oral History in Information Systems Research: A Reconsideration of a Traditional Tool. In: Proceedings of the 6th Asian Conference on Information Systems. 2017 Presented at: ACIS'17; December 12-14, 2017; Phnom Penh, Cambodia URL: https://pdfs.semanticscholar.org/ddc6/5b58451ee87ba2b9747be4d361e77228cfd3.pdf?_ga=2.171117287.698584668.1584954169-1679671381.1567599385
21. Sarnecky MT. Historiography: a legitimate research methodology for nursing. ANS Adv Nurs Sci 1990 Jul;12(4):1-10. [Medline: 2112886]
22. Lewenson SB. Doing historical research. In: Lewenson SB, Herrmann EK, editors. Capturing Nursing History: A Guide to Historical Methods in Research. New York, USA: Springer Publishing; 2008:25-44.
23. Boschma G. Conducting oral history research in community mental health nursing. In: de Chesnay M, editor. Nursing Research Using Historical Methods: Qualitative Designs and Methods in Nursing. New York, USA: Springer Publishing; 2015:85-104.
24. Alberta Netcare. 2015 Apr. An Overview of Alberta's Electronic Health Record Information System URL: http://www.albertanetcare.ca/documents/An_Overview_of_Albertas_ERHIS.pdf [accessed 2018-02-05]
25. Graham TA, Ballermann M, Lang E, Bullard MJ, Parsons D, Mercuur G, et al. Emergency physician use of the Alberta Netcare portal, a province-wide interoperable electronic health record: multi-method observational study. JMIR Med Inform 2018 Sep 25;6(3):e10184 [FREE Full text] [doi: 10.2196/10184] [Medline: 30274967]
26. Lee J. CBC News. 2019 Jan 29. New App Will Let Albertans See Their Own Health Records URL: https://www.cbc.ca/news/canada/calgary/alberta-health-information-phone-app-1.4995583 [accessed 2019-02-03]
27. Alberta Health Services. Connect Care URL: https://www.albertahealthservices.ca/info/cis.aspx [accessed 2019-04-16]
28. Walker DM, Sieck CJ, Menser T, Huerta TR, McAlearney AS. Information technology to support patient engagement: where do we stand and where can we go? J Am Med Inform Assoc 2017 Nov 1;24(6):1088-1094. [doi: 10.1093/jamia/ocx043] [Medline: 28460042]
29. Born K, Sullivan T, Bear R. Healthy Debate. 2013 Oct 13. Restructuring Alberta's Health System URL: https://healthydebate.ca/2013/10/topic/politics-of-health-care/restructuring-alberta-health [accessed 2019-05-10]
30. Church J, Smith N. Health reform in Alberta: Fiscal crisis, political leadership, and institutional change within a single-party democratic state. In: Lazar H, Forest PG, Lavis JN, Church J, editors. Paradigm Freeze: Why It Is So Hard to Reform Health Care in Canada. Robert Sutherland Hall: Queen's Policy Studies Series - School; 2014:478-480.
31. Kooij L, Groen WG, van Harten WH. Barriers and facilitators affecting patient portal implementation from an organizational perspective: qualitative study. J Med Internet Res 2018 May 11;20(5):e183 [FREE Full text] [doi: 10.2196/jmir.8989] [Medline: 29752253]
32. Koivunen M, Hätönen H, Välimäki M. Barriers and facilitators influencing the implementation of an interactive internet-portal application for patient education in psychiatric hospitals. Patient Educ Couns 2008 Mar;70(3):412-419. [doi: 10.1016/j.pec.2007.11.002] [Medline: 18079085]
33. McGinn CA, Grenier S, Duplantie J, Shaw N, Sicotte C, Mathieu L, et al. Comparison of user groups' perspectives of barriers and facilitators to implementing electronic health records: a systematic review. BMC Med 2011 Apr 28;9:46 [FREE Full text] [doi: 10.1186/1741-7015-9-46] [Medline: 21524315]
34. Richer M, Marchionni C, Lavoie-Tremblay M, Aubry M. The project management office: transforming healthcare in the context of a hospital redevelopment project. Healthc Manage Forum 2013;26(3):150-156. [doi: 10.1016/j.hcmf.2013.05.001] [Medline: 24409583]
35. Aubry M, Richer M, Lavoie-Tremblay M. Governance performance in complex environment: The case of a major transformation in a university hospital. Int J Proj Manage 2014 Nov;32(8):1333-1345. [doi: 10.1016/j.ijproman.2013.07.008]
36. HealthIT. How Do I Select a Vendor? URL: https://www.healthit.gov/faq/how-do-i-select-vendor [accessed 2019-07-15]

37. McDowell SW, Wahl R, Michelson J. Herding cats: the challenges of EMR vendor selection. J Healthc Inf Manag 2003;17(3):63-71. [Medline: 12858600]
38. Middleton B. Evaluating the value of healthcare information technology: finding the diamond in the rough, and tumble. AMIA Annu Symp Proc 2006;2006:1172-1173 [FREE Full text]
39. Hazara AM, Bhandari S. Barriers to patient participation in a self-management and education website Renal PatientView: A questionnaire-based study of inactive users. Int J Med Inform 2016 Mar;87:10-14. [doi: 10.1016/j.ijmedinf.2015.12.004] [Medline: 26806707]
40. Rudin RS, Jones SS, Shekelle P, Hillestad RJ, Keeler EB. The value of health information technology: filling the knowledge gap. Am J Manag Care 2014 Nov;20(11 Spec No. 17):eSP1-eSP8 [FREE Full text] [Medline: 25811814]
41. Goldzweig CL, Orshansky G, Paige NM, Towfigh AA, Haggstrom DA, Miake-Lye I, et al. Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. Ann Intern Med 2013 Nov 19;159(10):677-687. [doi: 10.7326/0003-4819-159-10-201311190-00006] [Medline: 24247673]
42. Buntin MB, 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. [doi: 10.1377/hlthaff.2011.0178] [Medline: 21383365]
43. Bata S, Richardson T. Value of investment as a key driver for prioritization and implementation of healthcare software. Perspect Health Inf Manag 2018;15(Winter):1g [FREE Full text] [Medline: 29618963]
44. Meyer R, Degoulet P. Choosing the right amount of healthcare information technologies investments. Int J Med Inform 2010 Apr;79(4):225-231. [doi: 10.1016/j.ijmedinf.2010.01.001] [Medline: 20122867]
45. Mamlin BW, Tierney WM. The promise of information and communication technology in healthcare: extracting value from the chaos. Am J Med Sci 2016 Jan;351(1):59-68. [doi: 10.1016/j.amjms.2015.10.015] [Medline: 26802759]
46. Wienert J. Understanding health information technologies as complex interventions with the need for thorough implementation and monitoring to sustain patient safety. Front ICT 2019 May 17;6:9. [doi: 10.3389/fict.2019.00009]
47. Cresswell KM, Bates DW, Sheikh A. Ten key considerations for the successful implementation and adoption of large-scale health information technology. J Am Med Inform Assoc 2013 Jun;20(e1):e9-13 [FREE Full text] [doi: 10.1136/amiajnl-2013-001684] [Medline: 23599226]
48. Jeffries M, Phipps DL, Howard RL, Avery AJ, Rodgers S, Ashcroft DM. Understanding the implementation and adoption of a technological intervention to improve medication safety in primary care: a realist evaluation. BMC Health Serv Res 2017 Mar 14;17(1):196 [FREE Full text] [doi: 10.1186/s12913-017-2131-5] [Medline: 28288634]
49. Otte-Trojel T, de Bont A, Rundall TG, van de Klundert J. What do we know about developing patient portals? a systematic literature review. J Am Med Inform Assoc 2016 Apr;23(e1):e162-e168 [FREE Full text] [doi: 10.1093/jamia/ocv114] [Medline: 26335985]
50. Dendere R, Slade C, Burton-Jones A, Sullivan C, Staib A, Janda M. Patient portals facilitating engagement with inpatient electronic medical records: a systematic review. J Med Internet Res 2019 Apr 11;21(4):e12779 [FREE Full text] [doi: 10.2196/12779] [Medline: 30973347]
51. Yen P, Walker DM, Smith JM, Zhou MP, Menser TL, McAlearney AS. Usability evaluation of a commercial inpatient portal. Int J Med Inform 2018 Feb;110:10-18. [doi: 10.1016/j.ijmedinf.2017.11.007] [Medline: 29331248]
52. Hefner JL, Sieck CJ, Walker DM, Huerta TR, McAlearney AS. System-wide inpatient portal implementation: survey of health care team perceptions. JMIR Med Inform 2017 Sep 14;5(3):e31 [FREE Full text] [doi: 10.2196/medinform.7707] [Medline: 28912115]
53. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. J Am Med Inform Assoc 2006;13(2):121-126 [FREE Full text] [doi: 10.1197/jamia.M2025] [Medline: 16357345]
54. Cline GB, Luiz JM. Information technology systems in public sector health facilities in developing countries: the case of South Africa. BMC Med Inform Decis Mak 2013 Jan 24;13:13 [FREE Full text] [doi: 10.1186/1472-6947-13-13] [Medline: 23347433]
55. Kaelber D, Pan E. The value of personal health record (PHR) systems. AMIA Annu Symp Proc 2008 Nov 6:343-347 [FREE Full text] [Medline: 18999276]
56. Powell K, Myers C. Electronic patient portals: patient and provider perceptions. Online J Nurs Inform 2018 Feb;22(1):- [FREE Full text]
57. Baldwin JL, Singh H, Sittig DF, Giardina TD. Patient portals and health apps: pitfalls, promises, and what one might learn from the other. Healthc (Amst) 2017 Sep;5(3):81-85 [FREE Full text] [doi: 10.1016/j.hjdsi.2016.08.004] [Medline: 27720139]
58. Powell KR. Patient-perceived facilitators of and barriers to electronic portal use: a systematic review. Comput Inform Nurs 2017 Nov;35(11):565-573. [doi: 10.1097/CIN.0000000000000377] [Medline: 28723832]
59. Kipping S, Stuckey MI, Hernandez A, Nguyen T, Riahi S. A web-based patient portal for mental health care: benefits evaluation. J Med Internet Res 2016 Nov 16;18(11):e294 [FREE Full text] [doi: 10.2196/jmir.6483] [Medline: 27852556]
60. Titler MG. The evidence for evidence-based practice implementation. In: Huges RG, editor. Patient Safety and Quality: An Evidence-Based Handbook for Nurses. Rockville, MD: Dept of Health and Human Services; 2008.

61. Houston TK, Richardson LM, Cotten SR. Patient-directed digital health technologies: is implementation outpacing evidence? Med Care 2019 Feb;57(2):95-97. [doi: 10.1097/MLR.0000000000001068] [Medline: 30601341]
62. Knaup P, Bott O, Kohl C, Lovis C, Garde S. Electronic patient records: moving from islands and bridges towards electronic health records for continuity of care. Yearb Med Inform 2007:34-46. [Medline: 17700902]
63. Rigby M, Georgiou A, Hyppönen H, Ammenwerth E, de Keizer N, Magrabi F, et al. Patient portals as a means of information and communication technology support to patient- centric care coordination - the missing evidence and the challenges of evaluation. A joint contribution of IMIA WG EVAL and EFMI WG EVAL. Yearb Med Inform 2015 Aug 13;10(1):148-159 [FREE Full text] [doi: 10.15265/IY-2015-007] [Medline: 26123909]

Abbreviations

AH: Alberta Health
AHS: Alberta Health Services
ANP: Alberta Netcare Portal
EHR: electronic health record
EMR: electronic medical record
IT: information technology
PCNs: primary care networks
PHR: personal health record
PoC: proof of concept
RFPs: request for proposals

Edited by G Eysenbach; submitted 17.12.19; peer-reviewed by N Shaw, C Slade; comments to author 15.01.20; revised version received 10.02.20; accepted 27.02.20; published 26.05.20.

Please cite as:
Avdagovska M, Stafinski T, Ballermann M, Menon D, Olson K, Paul P
Tracing the Decisions That Shaped the Development of MyChart, an Electronic Patient Portal in Alberta, Canada: Historical Research Study
J Med Internet Res 2020;22(5):e17505
URL: http://www.jmir.org/2020/5/e17505/
doi:10.2196/17505
PMID:32452811

Original Paper

Willingness of Patients Prescribed Medications for Lifestyle-Related Diseases to Use Personal Health Records: Questionnaire Study

Ryoma Hirano[1,2*], MS; Satoko Yamaguchi[2,3*], MD, PhD; Kayo Waki[2,4,5], MD, PhD; Yoshihiko Kimura[1], BS; Keiichi Chin[1], PharmD; Yasuhito Nannya[6], MD, PhD; Masaomi Nangaku[7], MD, PhD; Takashi Kadowaki[3,5], MD, PhD; Kazuhiko Ohe[4], MD, PhD

[1]Nihon Chouzai Co, Ltd, Tokyo, Japan

[2]Department of Ubiquitous Health Informatics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

[3]Department of Prevention of Diabetes and Lifestyle-Related Diseases, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

[4]Department of Biomedical Informatics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

[5]Department of Diabetes and Metabolic Diseases, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

[6]Department of Pathology and Tumor Biology, Kyoto University, Kyoto, Japan

[7]Division of Nephrology and Endocrinology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

[*]these authors contributed equally

Corresponding Author:
Kayo Waki, MD, PhD
Department of Biomedical Informatics
Graduate School of Medicine
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku
Tokyo, 113-8655
Japan
Phone: 81 3 5800 9129
Fax: 81 3 5800 9129
Email: kwaki-tky@umin.ac.jp

Abstract

Background: Personal health record (PHR) systems let individuals utilize their own health information to maintain and improve quality of life. Using PHRs is expected to support self-management in patients with lifestyle-related diseases.

Objective: The aim of this study was to identify predictors of the willingness to use PHRs among patients who are prescribed medications for lifestyle-related diseases.

Methods: We recruited pharmacy patrons, aged 20 years or older, who had received at least one medication indicated for hypertension, dyslipidemia, or diabetes. Participants completed self-administered questionnaires regarding their previous diseases, awareness of health care, experience in using PHRs, willingness to use PHRs, and barriers to using PHRs. Data were analyzed using multivariate logistic regression models.

Results: Of the 3708 subjects meeting eligibility criteria, 2307 replies (62.22%) were collected. While only 174 (7.54%) participants had previous PHR experience, 853 (36.97%) expressed willingness to use PHRs. In the multivariate analysis, *considering exercise to be important for health management* (odds ratio [OR] 1.57, 95% CI 1.12-2.21; *P*=.009), *obtaining medical information from books or magazines* (OR 1.23, 95% CI 0.96-1.59; *P*=.10), and *obtaining medical information from the internet* (OR 1.45, 95% CI 1.13-1.87; *P*=.004) were newly identified predictors. These were in addition to known predictors, such as being employed, owning information terminals, and previous PHR experience.

Conclusions: Patients who have an active and positive attitude toward health seem to be more willing to use PHRs. Investigating willingness should contribute to the development of more useful PHRs for self-management among patients prescribed medications for lifestyle-related diseases.

(J Med Internet Res 2020;22(5):e13866) doi:10.2196/13866

KEYWORDS

personal health records; chronic disease; health behavior; eHealth; mHealth; mobile phone

Introduction

Hypertension, type 2 diabetes, dyslipidemia, and obesity are referred to as lifestyle-related diseases and are major risk factors for coronary heart diseases and strokes if left untreated. The goal of treating lifestyle-related diseases is to maintain good control for a long period of time, staving off mortality. With lifestyle-related diseases, improving diet and exercise and maintaining good medication adherence are crucial [1,2] but doing so for a long period of time can be hard. Adherence to diabetic diet therapy and regular exercise was reported in one study to be 69% and 19%, respectively [3]. For patients with lifestyle-related diseases in general, low medication adherence has been reported [4-6].

Recently, tools that enhance information and communication technology (ICT) have garnered much attention for effectively supporting self-management of lifestyle-related diseases [7]. Prominent in such technology are personal health record (PHR) systems, which consist of information coordination that allows patients to manage and apply their personal health, medical care, and nursing information. The use of PHRs is expected to help prevent progression of lifestyle-related diseases by improving patients' lifestyles and medication adherence [8,9]. The reported benefits of PHRs were engaging patients in their health behavior, improving their ability to self-manage, and improving communication with health care providers [10,11]. Self-monitoring blood pressure or blood glucose level by using PHRs makes it possible to give patients more specific and direct feedback [10]. A number of PHR systems were developed to support self-management in diabetes patients and were shown to be effective for improving glycemic control [12-14].

We had previously investigated the willingness of diabetes patients to use ICT-based self-management tools and identified predictors of willingness, such as current use of ICT devices or previous dropout from diabetic care [15,16]. However, the role of awareness of health management or medication adherence in this process was not understood, especially among the older population, which accounts for a large percentage of patients with lifestyle-related diseases. Recently, the importance of understanding technology-based behavior in the older population has been emphasized [17,18]. However, little has been understood about willingness and barriers to using PHRs for self-management in the older population. Therefore, we conducted an exploratory study that investigated willingness to use PHRs and awareness of health management in a larger population of patients prescribed medications for lifestyle-related diseases and identified predictors of willingness. Understanding the predictors of willingness to use PHRs may help identify the target population and develop more user-friendly PHRs that could be useful for self-management.

Methods

Study Design and Participants

Between February 6 and March 4, 2017, patients who visited any of the 85 Nihon Chouzai pharmacies in Kanagawa Prefecture, Japan, were recruited for the study. Eligibility criteria were being 20 years of age or older and having been prescribed at least one medication typically associated with treatment for hypertension, dyslipidemia, or diabetes (ie, for hypertension, receiving calcium channel blockers, angiotensin II receptor blockers, angiotensin-converting enzyme inhibitors, diuretics, beta adrenergic blockers, or alpha-beta adrenergic blockers; for dyslipidemia, receiving 3-hydroxyl-3-methylglutaryl-CoA reductase inhibitors [ie, statins], fibrate, anion exchange resin, Niemann-Pick C1-Like 1 inhibitor, probucol, nicotinate, or eicosapentaenoic acid; and for diabetes, receiving biguanide, thiazolidine derivatives, sulfonylurea, glinide, dipeptidyl peptidase-4 inhibitor, alfa-glucosidase inhibitor, sodium glucose cotransporter 2 inhibitor, insulin, or glucagon-like peptide-1 receptor agonist [19]). All participants provided written informed consent.

Exclusion criteria were inability to communicate, visiting the pharmacy on behalf of someone else, being judged by pharmacists as inappropriate, receiving in-home treatment, or being admitted to a nursing home.

Of those scheduled to visit the pharmacy during the 4-week study period, 29,771 who met the eligibility criteria were targeted. To assure the collection of questionnaires from 10% of the patients, 5000 were preselected randomly, positing an estimated response rate of 60%. The number of preselected patients at each pharmacy was in proportion to the number of patients who had visited a corresponding pharmacy during the 3 months immediately prior to the study period.

The targeted patients were given both a verbal and written explanation of the study; PHRs were explained both verbally and in writing as "systems such as electronic prescription records or health management apps to manage one's own medication and health information using electronic devices, including personal computers or smartphones." Those who responded with written informed consent were asked to complete a self-administered questionnaire. The pharmacists checked questionnaires for completion and collected them.

Questionnaire Contents

Each questionnaire elicited the following: (1) basic data, including gender, age, and occupation; (2) medical history, including the disease currently being treated, hospitalizations (with time of stay), and diagnosis of metabolic syndrome; (3) awareness of health management, including what patients considered important for maintaining health, their health information sources, the number of regularly prescribed medications, number of doses, their understanding of their diseases, use of pharmacies, and medication adherence; and (4) PHR-related data, including whether they own an information

terminal, their use of PHRs, and their willingness to use a PHR (see Multimedia Appendix 1). All information including medical history was self-reported and diagnoses were not validated by medical records.

Statistical Analyses

The differences between the two groups determined by their willingness to use PHRs were tested for statistical significance using the Fisher exact test. Correlation coefficients between two variables were calculated by the Spearman rank correlation test and were confirmed to be <.8 in all combinations.

The variables were analyzed using univariate logistic regression models. Those variables with an alpha significance level of .10 were considered candidates for explanatory variables in multivariate analyses.

Multivariate logistic regression models were built by stepwise model selection using Akaike's information criterion. The multivariate analyses included 2020 participants with no missing data in the variables that would comprise the final models.

All statistical analyses were carried out using EZR (Easy R) (Saitama Medical Center, Jichi Medical University, Saitama, Japan) [20].

Data Exclusion

A few questions on the questionnaire were excluded from the analyses because participants presumably misunderstood them; for example, "use of the family pharmacist system" and "use of Okusuri Techo Plus," which is an electronic prescription record application for smartphones and personal computers developed by Nihon Chouzai that has various functions, such as sending prescription information to the pharmacy, recording test results, management of medication adherence, and health information distribution.

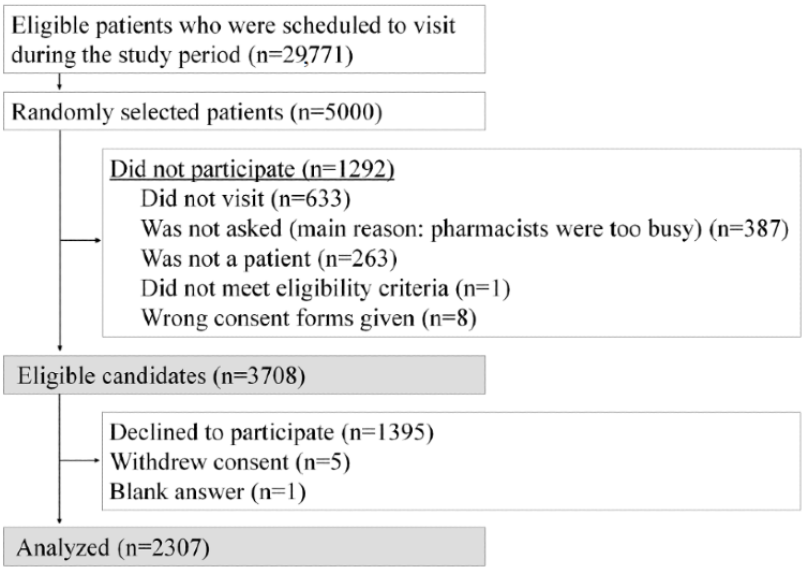
Although only 20.66% (1033/5000) of the preselected candidates had registered with a family pharmacist, 43.43% (1002/2307) of the participants responded that they used the family pharmacist system. Similarly, although the actual usage ratio of Okusuri Techo Plus stood at only 5.92% (296/5000) among preselected candidates, 45.08% (1040/2307) answered that they used it.

Results

Study Population

As described in the Methods section, 5000 patients were preselected. A total of 633 people did not visit the pharmacies with preselected patients' prescriptions during the study period. In total, 387 people were not given an explanation about the study, mainly because pharmacists were too busy, and 263 were excluded because they came to collect someone else's medication. Among the patients who gave written informed consent, 9 were excluded: 1 was not of age and 8 were given the wrong consent forms by their pharmacists. Overall, 3708 were considered eligible candidates. In total, 1395 declined to participate, 5 withdrew from the study before completing the questionnaire, and 1 turned in a blank answer form, leaving a final number of 2307 valid responses. That made the valid response rate 62.22% (2307/3708) (see Figure 1).

Figure 1. Flowchart of participants in this study. Random selection from a population of patients prescribed medications for lifestyle-related diseases, confirmation of eligibility, informed consent, and questionnaire collection.



As shown in Table 1, out of 2307 respondents, 1269 (55.01%) were male, 1636 (70.91%) were 60 years of age or older, and 924 (40.05%) were employed; 1345 (58.30%) reported they had hypertension, 705 (30.56%) had diabetes, and 432 (18.73%) had dyslipidemia. Out of 2307 respondents, 345 (14.95%) who reported neither hypertension, diabetes, nor dyslipidemia were included in the analysis, though some of them may have been treated for other diseases using the same medications used to treat hypertension, diabetes, and hyperlipidemia. Out of 2307 respondents, 1302 (56.44%) owned either a personal computer, smartphone, or tablet, with 899 out of 2307 (38.97%) having a smartphone. Out of 2307 respondents, 174 (7.54%) answered that they had experience using a PHR and 853 (36.97%) expressed willingness to use a PHR.

Table 1. Univariate models of willingness to use personal health records (PHRs).

Characteristic	Total (N=2307), n (%)[a]	Willingness to use PHRs (n=2155), n (%)[b]		Odds ratio (95% CI)	*P* value
		Yes (n=853)	No (n=1302)		
Gender					
Male	1269 (55.01)	548 (64.2)	648 (49.77)	1.81 (1.51-2.17)	<.001
Female	1007 (43.65)	298 (34.9)	637 (48.92)	1.00	
Age (years)					
≥60	1636 (70.91)	488 (57.2)	1017 (78.11)	0.36 (0.30-0.44)	<.001
<60	654 (28.35)	363 (42.6)	274 (21.04)	1.00	
Employment status					
Employed	924 (40.05)	468 (54.9)	414 (31.80)	2.57 (2.15-3.09)	<.001
Not employed	1352 (58.60)	380 (44.5)	866 (66.51)	1.00	
Medical history					
Hypertension	1345 (58.30)	471 (55.2)	782 (60.06)	0.81 (0.67-0.97)	.02
Diabetes	705 (30.56)	276 (32.4)	382 (29.34)	1.15 (0.95-1.39)	.16
Dyslipidemia	432 (18.73)	191 (22.4)	225 (17.28)	1.37 (1.10-1.72)	.004
Angina pectoris or myocardial infarction	265 (11.49)	91 (10.7)	150 (11.52)	0.91 (0.68-1.21)	.53
Arrhythmia	205 (8.89)	77 (9.0)	115 (8.83)	1.02 (0.74-1.39)	.94
Kidney disease	142 (6.16)	52 (6.1)	78 (5.99)	1.01 (0.69-1.48)	>.99
Hyperuricemia	125 (5.42)	59 (6.9)	64 (4.92)	1.43 (0.98-2.10)	.06
Cerebrovascular disorder	132 (5.72)	40 (4.7)	81 (6.22)	0.74 (0.49-1.10)	.13
Hospitalization experience					
Yes	1426 (61.81)	515 (60.4)	814 (62.52)	0.86 (0.71-1.03)	.10
No	796 (34.50)	321 (37.6)	436 (33.49)	1.00	
Metabolic syndrome					
Yes	543 (23.54)	258 (30.2)	257 (19.74)	1.73 (1.40-2.13)	<.001
No	1506 (65.28)	514 (60.3)	886 (68.05)	1.00	
Unknown	239 (10.36)	78 (9.1)	145 (11.14)	0.93 (0.68-1.26)	
What patients considered important for their health management					
Exercise	1975 (85.61)	774 (90.7)	1081 (83.03)	1.98 (1.50-2.64)	<.001
Diet	2066 (89.55)	795 (93.2)	1146 (88.02)	1.84 (1.33-2.57)	<.001
Sleep	1593 (69.05)	628 (73.6)	867 (66.59)	1.39 (1.15-1.70)	<.001
Prescription drugs	1188 (51.50)	431 (50.5)	672 (51.61)	0.95 (0.80-1.14)	.60
Over-the-counter drugs	44 (1.91)	14 (1.6)	29 (2.23)	0.73 (0.35-1.44)	.43
Health food and supplements	200 (8.67)	64 (7.5)	121 (9.29)	0.79 (0.57-1.09)	.16
What patients were willing to incorporate into daily life in managing health					
Exercise	1451 (62.90)	554 (64.9)	795 (61.06)	1.19 (0.99-1.43)	.06
Healthy diet	1907 (82.66)	714 (83.7)	1071 (82.26)	1.12 (0.88-1.42)	.35
Sleep	1272 (55.14)	479 (56.2)	711 (54.61)	1.07 (0.89-1.28)	.48
Prescription drugs	1342 (58.17)	505 (59.2)	745 (57.22)	1.09 (0.91-1.30)	.35
Over-the-counter drugs	38 (1.65)	16 (1.9)	18 (1.38)	1.37 (0.65-2.85)	.38
Health food and supplements	254 (11.01)	94 (11.0)	145 (11.14)	0.99 (0.74-1.31)	>.99
Health information sources					
Doctors	1939 (84.05)	742 (87.0)	1078 (82.80)	1.37 (1.06-1.77)	.01

Characteristic	Total (N=2307), n (%)[a]	Willingness to use PHRs (n=2155), n (%)[b]		Odds ratio (95% CI)	*P* value
		Yes (n=853)	No (n=1302)		
Nurses	208 (9.02)	80 (9.4)	116 (8.91)	1.06 (0.77-1.44)	.76
Pharmacists	398 (17.25)	176 (20.6)	197 (15.13)	1.45 (1.15-1.83)	.001
Dieticians	97 (4.20)	45 (5.3)	48 (3.69)	1.45 (0.93-2.25)	.08
Acquaintances	255 (11.05)	98 (11.5)	143 (10.98)	1.05 (0.79-1.39)	.73
Family	311 (13.48)	128 (15.0)	164 (12.60)	1.22 (0.94-1.58)	.12
Books or magazines	454 (19.68)	194 (22.7)	233 (17.90)	1.35 (1.08-1.68)	.007
Newspapers	450 (19.51)	161 (18.9)	261 (20.05)	0.93 (0.74-1.16)	.51
Television	836 (36.24)	306 (35.9)	475 (36.48)	0.97 (0.81-1.17)	.75
Internet	722 (31.30)	432 (50.6)	266 (20.43)	3.98 (3.28-4.84)	<.001
Understanding of one's diseases					
Understanding one's diseases or symptoms well	2055 (89.08)	808 (94.7)	1137 (87.33)	2.61 (1.83-3.79)	<.001
Understanding treatment methods well	1472 (63.81)	592 (69.4)	805 (61.83)	1.39 (1.15-1.68)	<.001
Understanding how to deal with deterioration of health	591 (25.62)	249 (29.2)	312 (23.96)	1.30 (1.07-1.59)	.009
Entrusting doctors with disease and health-related matters	910 (39.45)	312 (36.6)	539 (41.40)	0.81 (0.68-0.97)	.02
Not paying much attention to one's disease	130 (5.64)	33 (3.9)	82 (6.30)	0.60 (0.38-0.91)	.01
Information terminals					
Personal computer	958 (41.53)	526 (61.7)	404 (31.03)	3.54 (2.94-4.27)	<.001
Smartphone	899 (38.97)	522 (61.2)	354 (27.19)	4.19 (3.47-5.07)	<.001
Basic cell phone	687 (29.78)	215 (25.2)	447 (34.33)	0.63 (0.52-0.77)	<.001
Tablet device	296 (12.83)	190 (22.3)	99 (7.60)	3.44 (2.63-4.51)	<.001
Number of regularly prescribed medications					
≥3	1670 (72.39)	592 (69.4)	969 (74.42)	0.75 (0.61-0.91)	.003
<3	597 (25.88)	258 (30.2)	315 (24.19)	1.00	
Number of doses per day					
≥2	1460 (63.29)	520 (61.0)	848 (65.13)	0.80 (0.67-0.96)	.02
<2	784 (33.98)	324 (38.0)	423 (32.49)	1.00	
Monthly visit frequency to a pharmacy					
≥2	419 (18.16)	160 (18.8)	225 (17.28)	1.10 (0.87-1.38)	.42
<2	1844 (79.93)	686 (80.4)	1060 (81.41)	1.00	
Medication adherence					
Good	1762 (76.38)	638 (74.8)	1032 (79.26)	0.76 (0.61-0.94)	.01
Poor	506 (21.93)	209 (24.5)	257 (19.74)	1.00	
Experience using a PHR					
Yes	174 (7.54)	164 (19.2)	10 (0.77)	30.30 (15.92-64.79)	<.001
No	2059 (89.25)	686 (80.4)	1269 (97.47)	1.00	

[a]Percentages are out of 2307; because of missing responses, the percentages may not add up to 100%.

[b]Percentages are out of 853 (for *Yes*) or 1302 (for *No*); because of missing responses, the percentages may not add up to 100%.

Univariate Analysis

Factors associated with willingness to use PHRs were identified by univariate analysis (see Table 1). Among the basic characteristics, being male (odds ratio [OR] 1.81, 95% CI 1.51-2.17; *P*<.001), and being employed (OR 2.57, 95% CI 2.15-3.09; *P*<.001) were positive predictive factors, while being 60 years of age or older (OR 0.36, 95% CI 0.30-0.44; *P*<.001) was a negative predictive factor.

Regarding medical history, having dyslipidemia (OR 1.37, 95% CI 1.10-1.72; *P*=.004) and having been diagnosed with metabolic syndrome (OR 1.73, 95% CI 1.40-2.13; *P*<.001) were positive predictive factors, while having hypertension (OR 0.81, 95% CI 0.67-0.97; *P*=.02) was a negative predictive factor.

Regarding what patients considered important for their health management, positive predictive factors were exercise (OR 1.98, 95% CI 1.50-2.64; *P*<.001), diet (OR 1.84, 95% CI 1.33-2.57; *P*<.001), and sleep (OR 1.39, 95% CI 1.15-1.70; *P*<.001). Other positive predictive factors were sources of medical information, such as doctors (OR 1.37, 95% CI 1.06-1.77; *P*=.01), pharmacists (OR 1.45, 95% CI 1.15-1.83; *P*=.001), books or magazines (OR 1.35, 95% CI 1.08-1.68; *P*=.007), and the internet (OR 3.98, 95% CI 3.28-4.84; *P*<.001), while nurses, dieticians, acquaintances, family, newspapers, and television were not. Negative predictive factors were *regularly taking three or more types of medicine* (OR 0.75, 95% CI 0.61-0.91; *P*=.003), *taking medicine twice or more per day* (OR 0.80, 95% CI 0.67-0.96; *P*=.02), and *good adherence to taking medicine* (OR 0.76, 95% CI 0.61-0.94; *P*=.01).

Regarding understanding of diseases, positive predictive factors were *understanding one's diseases or symptoms well* (OR 2.61, 95% CI 1.83-3.79; *P*<.001), *understanding treatment methods well* (OR 1.39, 95% CI 1.15-1.68; *P*<.001), and *understanding how to deal with deterioration of health* (OR 1.30, 95% CI 1.07-1.59; *P*=.009). In contrast, *entrusting doctors with disease and health-related matters* (OR 0.81, 95% CI 0.68-0.97; *P*=.02) or *not paying much attention to one's disease* (OR 0.60, 95% CI 0.38-0.91; *P*=.01) were negative predictive factors.

Regarding PHR-related factors, owning a personal computer (OR 3.54, 95% CI 2.94-4.27; *P*<.001), smartphone (OR 4.19, 95% CI 3.47-5.07; *P*<.001), or tablet device (OR 3.44, 95% CI 2.63-4.51; *P*<.001) were positive predictors, while owning a basic cell phone (OR 0.63, 95% CI 0.52-0.77; *P*<.001) was a negative predictor. Having experience using a PHR (OR 30.30, 95% CI 15.92-64.79; *P*<.001) was a strong positive predictive factor.

Multivariate Logistic Regression Analyses

Multivariate logistic regression analyses were performed. The final model was identified as male (OR 1.52, 95% CI 1.22-1.90; *P*<.001), employed (OR 1.30, 95% CI 1.03-1.64; *P*=.03); having dyslipidemia (OR 1.25, 95% CI 0.97-1.62; *P*=.09); considering exercise important for health management (OR 1.57, 95% CI 1.12-2.21; *P*=.009); obtaining medical information from doctors (OR 1.36, 95% CI 1.01-1.83; *P*=.04), pharmacists (OR 1.63, 95% CI 1.24-2.13; *P*<.001), books or magazines (OR 1.23, 95% CI 0.96-1.59; *P*=.10), or the internet (OR 1.45, 95% CI 1.13-1.87; *P*=.004); understanding one's diseases or symptoms well (OR 1.37, 95% CI 0.92-2.05; *P*=.13); entrusting doctors with disease and health-related matters (OR 1.21, 95% CI 0.98-1.51; *P*=.08); owning a personal computer (OR 1.85, 95% CI 1.46-2.35; *P*<.001), smartphone (OR 2.04, 95% CI 1.61-2.59; *P*<.001), or tablet device (OR 1.36, 95% CI 0.99-1.86; *P*=.06); and having experience using a PHR (OR 16.60, 95% CI 8.52-32.20; *P*<.001). All these were positive predictors for the willingness to use a PHR (see Table 2).

Table 2. Univariate and multivariate logistic models of willingness to use personal health records (PHRs).

Characteristic	Univariate model (N=2307)		Multivariate model (n=2020)	
	Odds ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value
(Intercept)			0.06 (0.03-0.10)	<.001
Gender (male vs female)	1.81 (1.51-2.16)	<.001	1.52 (1.22-1.90)	<.001
Age (≥60 years vs <60)	0.36 (0.30-0.44)	<.001	N/A[a]	
Employment status (employed vs not employed)	2.58 (2.15-3.08)	<.001	1.30 (1.03-1.64)	.03
Medical history (yes vs no)				
Hypertension	0.81 (0.68-0.96)	.02	N/A[a]	
Diabetes	1.15 (0.95-1.38)	.15	N/A[b]	
Dyslipidemia	1.38 (1.11-1.71)	.004	1.25 (0.97-1.62)	.09
Angina pectoris or myocardial infarction	0.91 (0.69-1.20)	.52	N/A[b]	
Arrhythmia	1.02 (0.75-1.38)	.90	N/A[b]	
Kidney disease	1.01 (0.71-1.46)	.94	N/A[b]	
Hyperuricemia	1.43 (0.99-2.06)	.054	N/A[a]	
Cerebrovascular disorder	0.74 (0.50-1.09)	.13	N/A[b]	
Hospitalization experience	0.86 (0.72-1.03)	.10	N/A[b]	
Metabolic syndrome	1.73 (1.41-2.12)	<.001	N/A[a]	
What patients considered important for their health management (yes vs no)				
Exercise	1.98 (1.51-2.61)	<.001	1.57 (1.12-2.21)	.009
Diet	1.84 (1.34-2.52)	<.001	N/A[a]	
Sleep	1.39 (1.15-1.69)	<.001	N/A[a]	
Prescription drugs	0.95 (0.80-1.13)	.60	N/A[b]	
Over-the-counter drugs	0.73 (0.38-1.39)	.34	N/A[b]	
Health food and supplements	0.79 (0.58-1.08)	.15	N/A[b]	
What patients were willing to incorporate into daily life in managing health (yes vs no)				
Exercise	1.19 (0.99-1.42)	.06	N/A[a]	
Diet	1.12 (0.89-1.41)	.34	N/A[b]	
Sleep	1.07 (0.90-1.27)	.46	N/A[b]	
Prescription drugs	1.09 (0.91-1.30)	.34	N/A[b]	
Over-the-counter drugs	1.37 (0.69-2.69)	.37	N/A[b]	
Health food and supplements	0.99 (0.75-1.30)	.94	N/A[b]	
Health information sources (yes vs no)				
Doctors	1.37 (1.07-1.75)	.01	1.36 (1.01-1.83)	.04
Nurses	1.06 (0.78-1.42)	.72	N/A[b]	
Pharmacists	1.45 (1.16-1.82)	.001	1.63 (1.24-2.13)	<.001
Dieticians	1.45 (0.96-2.20)	.08	N/A[a]	
Acquaintances	1.05 (0.80-1.38)	.73	N/A[b]	
Family	1.22 (0.95-1.57)	.12	N/A[b]	

Characteristic	Univariate model (N=2307)		Multivariate model (n=2020)	
	Odds ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value
Books or magazines	1.35 (1.09-1.67)	.006	1.23 (0.96-1.59)	.10
Newspapers	0.93 (0.74-1.15)	.49	N/A[b]	
Television	0.97 (0.81-1.16)	.74	N/A[b]	
Internet	3.98 (3.29-4.82)	<.001	1.45 (1.13-1.87)	.004
Understanding of one's diseases (yes vs no)				
Understanding one's diseases or symptoms well	2.61 (1.84-3.70)	<.001	1.37 (0.92-2.05)	.13
Understanding treatment methods well	1.39 (1.16-1.67)	<.001	N/A[a]	
Understanding how to deal with deterioration of health	1.30 (1.07-1.58)	.008	N/A[a]	
Entrusting doctors with disease and health-related matters	0.81 (0.68-0.97)	.02	1.21 (0.98-1.51)	.08
Not paying much attention to one's disease	0.60 (0.40-0.90)	.01	N/A[a]	
Information terminals (yes vs no)				
Personal computer	3.54 (2.95-4.25)	<.001	1.85 (1.46-2.35)	<.001
Smartphone	4.19 (3.48-5.04)	<.001	2.04 (1.61-2.59)	<.001
Basic cell phone	0.63 (0.52-0.77)	<.001	N/A[a]	
Tablet device	3.44 (2.65-4.47)	<.001	1.36 (0.99-1.86)	.06
Number of regularly prescribed medications (≥3 vs <3)	0.75 (0.62-0.91)	.003	N/A[a]	
Number of doses per day (≥2 vs <2)	0.80 (0.67-0.96)	.02	N/A[a]	
Monthly visit frequency to a pharmacy (≥2 vs <2)	1.10 (0.88-1.38)	.41	N/A[b]	
Medication adherence (good vs poor)	0.76 (0.62-0.94)	.010	N/A[a]	
Experience using a PHR (yes vs no)	30.30 (15.90-57.80)	<.001	16.60 (8.52-32.20)	<.001

[a]N/A: not applicable; variables were considered candidates in the multivariate analyses (*P*<.10) but were discarded from the final model.

[b]N/A: not applicable; variables were not considered candidates for explanatory variables in multivariate analyses (*P*>.10).

Experience and Willingness to Use Personal Health Records

Participants who answered that they had used a PHR (174/2307, 7.54%) were asked the following question about their experience: Which applications had they used? The most common responses were *prescription records* (129/174, 74.1%), *activity amount recorders* (43/174, 24.7%), *blood pressure management* (37/174, 21.3%), and *weight management* (27/174, 15.5%) (see Table 3).

Table 3. Type of personal health record (PHR) application used by participants.

Type of application	Participants (n=174), n (%)
Prescription records	129 (74.1)
Activity amount recorders	43 (24.7)
Blood pressure management	37 (21.3)
Weight management	27 (15.5)
Information on sleep	13 (7.5)
Glycemic control	10 (5.7)
Menstrual management	4 (2.3)
Others	3 (1.7)

The types of information that respondents had most commonly managed through the use of a PHR included *medication prescription details* (76/174, 43.7%), *medication adherence* (54/174, 31.0%), and *home blood pressure values and body weight, etc* (50/174, 28.7%). When asked what types of information they wanted to manage, many responded with answers related to medication, such as *medication prescription details* (84/174, 48.3%) and *medication adherence* (75/174,

43.1%), as well as those related to test values, such as *test results at medical institution* (82/174, 47.1%) and *home blood pressure values and body weight, etc* (76/174, 43.7%) (see Table 4). It was found that while many people expressed a desire to manage test results at medical institutes using a PHR, they were not actually doing so (82/174, 47.1%, vs 20/174, 11.5%; *P*<.001, Fisher exact test). We also found a gap between participants' expressed desires and reality as to *home blood pressure values and body weight etc*, *medication adherence*, *record of the date of hospital visit*, *medical history*, and *meal content* (see Table 4). The most common sources of information about PHRs were *pharmacy staff* (107/174, 61.5%), followed by *the internet* (53/174, 30.5%) (see Table 5).

Table 4. Type of information that personal health record (PHR) users wanted to manage, or did manage, by PHR.

Information type	Participants (n=174), n (%)		*P* value
	Wanted to manage	Managed	
Medication prescription details	84 (48.3)	76 (43.7)	.45
Test results at medical institution	82 (47.1)	20 (11.5)	<.001
Home blood pressure values and body weight, etc	76 (43.7)	50 (28.7)	.005
Medication adherence	75 (43.1)	54 (31.0)	.03
Record of the date of hospital visit	40 (23.0)	21 (12.1)	.01
Medical history	35 (20.1)	8 (4.6)	<.001
Meal content	24 (13.8)	9 (5.2)	.010
Record of immunization	10 (5.7)	3 (1.7)	.09
Surgery record	6 (3.4)	2 (1.1)	.28
Medical history of the family	4 (2.3)	2 (1.1)	.68
Others	3 (1.7)	3 (1.7)	>.99

Table 5. Sources of information about personal health records (PHRs).

Information source[a]	Participants (n=174), n (%)
Pharmacy staff	107 (61.5)
Internet	53 (30.5)
Family	13 (7.5)
Poster	8 (4.6)
Television	6 (3.4)
Hospital staff	5 (2.9)
Newspapers	2 (1.1)
Books or magazines	2 (1.1)
Acquaintance	1 (0.6)
Others	8 (4.6)

[a]Participants who had previous PHR experience were asked where they got information from; multiple answers were allowed.

The 1302 participants who expressed unwillingness to use PHRs were asked why. The main reasons were *it needs more time and effort* (509/1302, 39.09%) or *the need is not felt* (465/1302, 35.71%), followed by *concern over security* (199/1302, 15.28%) (see Table 6).

Table 6. Reasons for unwillingness to use a personal health record (PHR).

Reasons[a]	Participants (n=1302), n (%)
More time and effort	509 (39.09)
The need is not felt	465 (35.71)
Concern over security	199 (15.28)
Cannot use mobile devices	35 (2.69)
Have no mobile device	17 (1.31)
Other	86 (6.61)
No answer	131 (10.06)

[a]Respondents who expressed unwillingness to use PHRs were asked the reasons why; multiple answers were allowed.

Discussion

Principal Findings

Among the predictive factors identified in this study, owning a personal computer, smartphone, or tablet and having experience using a PHR were consistent with prior reports that ICT literacy is associated with willingness to use a PHR [15,16,21-23]. While older age was reported to be a negative predictor in some studies [24,25], others reported that current use of ICT devices was more associated with willingness than age [21,23]. Consistently, in this study, although univariate analysis showed that older age was a negative predictor of willingness to use a PHR (see Table 1), that factor was discarded from the final model of the multivariate analysis (see Table 2). The ownership rate of personal computers, smartphones, and/or tablets was 88.7% (580/654) among participants under 60 years of age, but it was only 43.89% (718/1636) among participants 60 years of age or older (see Multimedia Appendix 2), clearly demonstrating the digital divide between generations. While 55.5% (363/654) of the participants under 60 years of age expressed willingness, only 29.83% (488/1636) of those 60 years of age or older did so (see Table 1). However, among participants 60 years of age or older, 48.1% (345/718) of those who owned ICT devices expressed willingness to use PHRs, compared to only 16.4% (136/831) of those who did not own these devices (see Multimedia Appendix 2). Being employed was identified as a predictor of willingness to use PHRs in this study, consistent with the previous reports [15,26].

While most of the previous reports suggested that gender was not a predictor [21,22,25,27], being male was identified as a predictor in this study. Consistently, the users of GlucoNote—an app for type 2 diabetes and prediabetes patients that we recently developed and released—were mostly men [28]. In contrast, studies in the United States and Europe reported that being female was a predictor of electronic health (eHealth) use [26,29]. These differences in results may be explained, at least partly, by different gender roles in each country; it was reported that a persistent gender gap in ICT use and skills was observed in Japan but not in the United States after adjusting for employment status [30].

The finding that participants who consider exercise important for their health management are more willing to use PHRs (see Table 2) was novel, as was the finding that patients who obtain medical information from doctors, pharmacists, books or magazines, and the internet are more willing to use PHRs. In contrast, obtaining medical information from newspapers or television was not associated with that willingness (see Table 2).

Regarding what the participants consider important for their health management, *diet*, *exercise*, and *sleep* were all positive predictors in the univariate analysis of willingness to use PHRs (see Table 1). However—and strikingly—only exercise remained a strong predictor in the multivariate analysis (OR 1.57, *P*=.009) (see Table 2), while diet and sleep were discarded from the final model. It was reported that for type 2 diabetes patients a health behavior change in physical activity was less common than with a change in healthy diet, and that changing motivation for a healthy diet was higher than that for habitual physical activity [31]. Exercise is different from eating and sleeping behavior in that one has to actively make an effort to implement it [32], so patients who value exercise may have a greater tendency to actively pursue health. Similarly, people who collect medical information from books, magazines, and the internet may be more active in pursuing health than those who obtain information from more passive media like newspapers and television. This is consistent with the previous report that a health-related information-seeking personality was a significant predictor for willingness to undergo online treatment [21].

Interestingly, while *considering exercise important for their health management* was a positive predictive factor, *is willing to incorporate exercise into daily life for health management* was not a significant predictor in the multivariate analysis in this study (see Table 2). According to the Transtheoretical Model, health behavior change consists of sequential stages: precontemplation, contemplation, preparation, action, and maintenance [33]. Although questionnaire choices in this study do not precisely correspond to the stages of change in the Transtheoretical Model, people who *consider exercise important for their health management* are thought to be in one of the contemplation, preparation, action, or maintenance stages, while people who are *willing to incorporate exercise into daily life for health management* are thought to be in the action or maintenance stage. Those who *consider exercise important for their health management* but are not *willing to incorporate exercise into daily life for health management* are likely to be in the contemplation or preparation stage. The result that

considering exercise important is a significant predictor of willingness to use PHRs—while *willing to incorporate exercise into daily life for health management* is not (see Table 2)—matches our previous report that diabetes patients who are in the contemplation or preparation stage of change in exercise are more willing to use ICT-based self-management tools [15]. Therefore, PHRs equipped with the function of stage-matched intervention for contemplation and preparation stages, such as understanding their own state, goal setting, or identification of barriers to behavior change, will lead the users to take action for health management, as stage-matched intervention was shown to be effective for improving glycemic control, hypertension control, and physical activity [34-37].

Because it was reported that health literacy was one of the key variables explaining a willingness to adopt a PHR [23], we expected that people who understand their disease well would be more willing to use a PHR. As expected, *understanding one's diseases or symptoms well*, *understanding treatment methods well*, and *understanding how to deal with deterioration of health* were all positive predictive factors for willingness to use PHRs, while *entrusting doctors with disease and health-related matters* and *not paying much attention to one's disease* were negative predictive factors in univariate analyses (see Table 1). In a multivariate analysis, however, *entrusting doctors with disease and health-related matters* was identified as a positive predictor, along with *understanding one's diseases or symptoms well*. The reason for this seemingly paradoxical result is that multiple answers were allowed: people who answered *I entrust doctors with disease and health-related matters* were also likely to have chosen *I understand treatment methods well* and *I understand how to deal with deterioration of health*, both of which were strong positive predictors in univariate analyses.

While people who had high health literacy were found to be more willing to adopt a PHR [23], those with a low health literacy level might benefit most from a PHR. Plainly, educating people about the benefits of using PHRs will be important.

People who had used PHRs were asked about their usage. Although 94.3% (164/174) of them were willing to use PHRs (see Table 1), gaps between the contemplated and actual use were found (see Table 4). While many people wished to manage information using a PHR, they were not doing so (see Table 4). Future tasks include developing a PHR equipped with functions desired by the users, with the hope that the additions will help trigger usage. The most frequently managed information using PHRs was prescription details, with which there was no gap between the desired and actual function (see Table 4); this may be due to the fact that this study was done in pharmacies, where staff encouraged patients to use electronic prescription records as described above.

As noted, our results showed that 1302 out of 2307 participants (56.44%) expressed no willingness to use a PHR (see Table 1). Consistent with previous reports [38], the main reason for this disinclination was that using PHRs requires too much time and effort (see Table 6). It might be possible to capture the interest of such holdouts if there was a function that made information management quick and easy—perhaps something like auto-linking data on the order of inspection values between patients and medical institutions.

Other frequently mentioned reasons for unwillingness to use PHRs include no need being felt and concern over security. This points to the need for further efforts to inform people of the merits of using PHRs, as well as improvements in security.

This study suggested that the most prevalent information source regarding PHRs was the internet, if we eliminate the potential bias of the fact that the study was conducted in pharmacies where pharmacists encouraged patients to use electronic prescription records (see Table 5). This means that only people with high ICT literacy were able to get enough information about PHRs. Obviously, the visibility and usability of PHRs must be improved so that even citizens with low ICT literacy will find it easy to use.

Limitations

There are several limitations to this study. First, because of the cross-sectional design of the study, the results do not reflect the long-term status of willingness to use PHRs, which might change drastically over time as the technology disseminates and comes into common use. Future tasks include updating the survey to follow the temporal changes in willingness.

Second, the survey was done in a geographically limited area: all the pharmacies were located in the Kanagawa Prefecture of Japan, considered part of the Greater Tokyo Area. Japan has the highest aging rate in the world with relatively high ICT literacy, so the results might not be valid for other regions with different age structures and levels of ICT literacy. Moreover, the eligibility criteria were determined by prescribed medications, which do not always correspond to a diagnosis. A previous study on claims-based algorithms for chronic conditions demonstrated that the sensitivity and specificity for identifying people with hypertension, diabetes, and dyslipidemia by using medication codes were 75.0% and 97.9%, 78.6% and 99.5%, and 34.6% and 97.2%, respectively [19]. Although the specificity is high, as expected, the sensitivity is relatively low, so the survey results cannot be generalized to the entire population with lifestyle-related diseases.

Third, because this was a self-administered questionnaire, some of the information such as the disease currently being treated may not be fully precise. As described above, 345 out of 2307 (14.95%) patients responded they were treated for neither hypertension, diabetes, nor dyslipidemia, even though all the participants were prescribed medication for at least one of the three diseases. Some of them may have been prescribed these medications to treat other diseases, but most of them presumably failed to self-report their diseases. Consistently, a previous report suggested that sensitivity of self-report questionnaires compared to pharmacy insurance claims was 92.4%, 82.6%, and 86.2% for hypertension, diabetes, and dyslipidemia, respectively [39]. Moreover, the intent of some questions may not have been fully understood or may have caused misunderstanding by the respondents, as it was not a structured interview. Importantly, although the concept of PHRs was explained both verbally and in writing as described in the Methods section, it is possible that the participants' insufficient

understanding or unfamiliarity with PHRs contributed to the lack of willingness, given that only 174 out of 2307 (7.54%) had experience using a PHR. Showing the participants actual examples of PHRs, such as health care management apps, prior to providing the questionnaire might help them better understand the concept of PHRs and properly determine the willingness to use PHRs.

Fourth, because the study was conducted at outlets of a dispensing pharmacy company whose staff encouraged patients to use electronic prescription records, there may have been biased results for some questions.

Conclusions

We identified factors predicting a willingness to use a PHR by patients prescribed medication for lifestyle-related diseases. Patients with an active attitude toward health, such as those who consider exercise important for health management or who obtain medical information from books, magazines, or the internet, seem to be more willing to use a PHR. The main barriers to using a PHR were time and effort required and lack of perceived necessity, followed by concern over security. Analyzing these predictors and barriers will prove useful in identifying suitable candidates as PHR users and in developing more helpful tools for self-management by patients with lifestyle-related diseases.

Acknowledgments

This study was funded by Nihon Chouzai Co, Ltd, and was conducted at the Department of Ubiquitous Health Informatics, which was engaged in a cooperative program between the University of Tokyo, NTT DOCOMO, Inc, and Nihon Chouzai Co, Ltd.

Conflicts of Interest

RH, YK, and KC are employees of Nihon Chouzai Co, Ltd. SY and TK are members of the Department of Prevention of Diabetes and Lifestyle-Related Diseases, which is part of a cooperative program between the University of Tokyo and Asahi Mutual Life Insurance Company. SY and KW were members of the Department of Ubiquitous Health Informatics when the study was conducted.

Multimedia Appendix 1
The questionnaire used in this study.
[DOCX File , 29 KB - jmir_v22i5e13866_app1.docx]

Multimedia Appendix 2
The ownership rate of information and communication technology (ICT) devices (ie, personal computer, smartphone, and/or tablet) and willingness to use personal health records (PHRs) by age group.
[DOCX File , 113 KB - jmir_v22i5e13866_app2.docx]

References

1. Patnode CD, Evans CV, Senger CA, Redmond N, Lin JS. Behavioral counseling to promote a healthful diet and physical activity for cardiovascular disease prevention in adults without known cardiovascular disease risk factors: Updated evidence report and systematic review for the US Preventive Services Task Force. JAMA 2017 Jul 11;318(2):175-193. [doi: 10.1001/jama.2017.3303] [Medline: 28697259]
2. Simpson SH, Eurich DT, Majumdar SR, Padwal RS, Tsuyuki RT, Varney J, et al. A meta-analysis of the association between adherence to drug therapy and mortality. BMJ 2006 Jul 01;333(7557):15 [FREE Full text] [doi: 10.1136/bmj.38875.675486.55] [Medline: 16790458]
3. Kravitz RL, Hays RD, Sherbourne CD, DiMatteo MR, Rogers WH, Ordway L, et al. Recall of recommendations and adherence to advice among patients with chronic medical conditions. Arch Intern Med 1993 Aug 23;153(16):1869-1878. [Medline: 8250648]
4. Abegaz TM, Shehab A, Gebreyohannes EA, Bhagavathula AS, Elnour AA. Nonadherence to antihypertensive drugs: A systematic review and meta-analysis. Medicine (Baltimore) 2017 Jan;96(4):e5641 [FREE Full text] [doi: 10.1097/MD.0000000000005641] [Medline: 28121920]
5. Cramer JA. A systematic review of adherence with medications for diabetes. Diabetes Care 2004 May;27(5):1218-1224. [doi: 10.2337/diacare.27.5.1218] [Medline: 15111553]
6. Tsuyuki RT, Bungard TJ. Poor adherence with hypolipidemic drugs: A lost opportunity. Pharmacotherapy 2001 May;21(5):576-582. [doi: 10.1592/phco.21.6.576.34541] [Medline: 11349746]
7. Ruiz Morilla MD, Sans M, Casasa A, Giménez N. Implementing technology in healthcare: Insights from physicians. BMC Med Inform Decis Mak 2017 Jun 27;17(1):92 [FREE Full text] [doi: 10.1186/s12911-017-0489-2] [Medline: 28655299]
8. Waki K, Fujita H, Uchimura Y, Omae K, Aramaki E, Kato S, et al. DialBetics: A novel smartphone-based self-management support system for type 2 diabetes patients. J Diabetes Sci Technol 2014 Mar;8(2):209-215 [FREE Full text] [doi: 10.1177/1932296814526495] [Medline: 24876569]

9. Bonoto BC, de Araújo VE, Godói IP, de Lemos LL, Godman B, Bennie M, et al. Efficacy of mobile apps to support the care of patients with diabetes mellitus: A systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2017 Mar 01;5(3):e4 [FREE Full text] [doi: 10.2196/mhealth.6309] [Medline: 28249834]
10. Price M, Bellwood P, Kitson N, Davies I, Weber J, Lau F. Conditions potentially sensitive to a personal health record (PHR) intervention: A systematic review. BMC Med Inform Decis Mak 2015 Apr 18;15:32 [FREE Full text] [doi: 10.1186/s12911-015-0159-1] [Medline: 25927384]
11. Afshin A, Babalola D, Mclean M, Yu Z, Ma W, Chen C, et al. Information technology and lifestyle: A systematic evaluation of internet and mobile interventions for improving diet, physical activity, obesity, tobacco, and alcohol use. J Am Heart Assoc 2016 Aug 31;5(9):e003058 [FREE Full text] [doi: 10.1161/JAHA.115.003058] [Medline: 27581172]
12. Yoshida Y, Boren SA, Soares J, Popescu M, Nielson SD, Simoes EJ. Effect of health information technologies on glycemic control among patients with type 2 diabetes. Curr Diab Rep 2018 Oct 18;18(12):130 [FREE Full text] [doi: 10.1007/s11892-018-1105-2] [Medline: 30338403]
13. Kitsiou S, Paré G, Jaana M, Gerber B. Effectiveness of mHealth interventions for patients with diabetes: An overview of systematic reviews. PLoS One 2017;12(3):e0173160 [FREE Full text] [doi: 10.1371/journal.pone.0173160] [Medline: 28249025]
14. Wu X, Guo X, Zhang Z. The efficacy of mobile phone apps for lifestyle modification in diabetes: Systematic review and meta-analysis. JMIR Mhealth Uhealth 2019 Jan 15;7(1):e12297 [FREE Full text] [doi: 10.2196/12297] [Medline: 30664494]
15. Shibuta T, Waki K, Tomizawa N, Igarashi A, Yamamoto-Mitani N, Yamaguchi S, et al. Willingness of patients with diabetes to use an ICT-based self-management tool: A cross-sectional study. BMJ Open Diabetes Res Care 2017;5(1):e000322 [FREE Full text] [doi: 10.1136/bmjdrc-2016-000322] [Medline: 28243450]
16. Yamaguchi S, Waki K, Tomizawa N, Waki H, Nannya Y, Nangaku M, et al. Previous dropout from diabetic care as a predictor of patients' willingness to use mobile applications for self-management: A cross-sectional study. J Diabetes Investig 2017 Jul;8(4):542-549 [FREE Full text] [doi: 10.1111/jdi.12613] [Medline: 28012247]
17. Haldane V, Tan YG, Teo KW, Koh JJ, Srivastava A, Cheng RX, et al. Perspectives on acceptance and use of a mobile health intervention for the prevention of atherosclerotic cardiovascular disease in Singapore: Mixed-methods study. JMIR Mhealth Uhealth 2019 Mar 14;7(3):e11108 [FREE Full text] [doi: 10.2196/11108] [Medline: 30869651]
18. Isaković M, Sedlar U, Volk M, Bešter J. Usability pitfalls of diabetes mHealth apps for the elderly. J Diabetes Res 2016;2016:1604609 [FREE Full text] [doi: 10.1155/2016/1604609] [Medline: 27034957]
19. Hara K, Tomio J, Svensson T, Ohkuma R, Svensson AK, Yamazaki T. Association measures of claims-based algorithms for common chronic conditions were assessed using regularly collected data in Japan. J Clin Epidemiol 2018 Jul;99:84-95. [doi: 10.1016/j.jclinepi.2018.03.004] [Medline: 29548842]
20. Kanda Y. Investigation of the freely available easy-to-use software 'EZR' for medical statistics. Bone Marrow Transplant 2013 Mar;48(3):452-458 [FREE Full text] [doi: 10.1038/bmt.2012.244] [Medline: 23208313]
21. Roettl J, Bidmon S, Terlutter R. What predicts patients' willingness to undergo online treatment and pay for online treatment? Results from a Web-based survey to investigate the changing patient-physician relationship. J Med Internet Res 2016 Feb 04;18(2):e32 [FREE Full text] [doi: 10.2196/jmir.5244] [Medline: 26846162]
22. Seidman Z, McNamara R, Wootton S, Leung R, Spencer L, Dale M, et al. People attending pulmonary rehabilitation demonstrate a substantial engagement with technology and willingness to use telerehabilitation: A survey. J Physiother 2017 Jul;63(3):175-181 [FREE Full text] [doi: 10.1016/j.jphys.2017.05.010] [Medline: 28652080]
23. Noblin AM, Wan TT, Fottler M. The impact of health literacy on a patient's decision to adopt a personal health record. Perspect Health Inf Manag 2012;9:1-13 [FREE Full text] [Medline: 23209454]
24. Serrano KJ, Yu M, Riley WT, Patel V, Hughes P, Marchesini K, et al. Willingness to exchange health information via mobile devices: Findings from a population-based survey. Ann Fam Med 2016;14(1):34-40 [FREE Full text] [doi: 10.1370/afm.1888] [Medline: 26755781]
25. Paslakis G, Fischer-Jacobs J, Pape L, Schiffer M, Gertges R, Tegtbur U, et al. Assessment of use and preferences regarding internet-based health care delivery: Cross-sectional questionnaire study. J Med Internet Res 2019 May 16;21(5):e12416 [FREE Full text] [doi: 10.2196/12416] [Medline: 31099338]
26. Torrent-Sellens J, Díaz-Chao Á, Soler-Ramos I, Saigí-Rubió F. Modelling and predicting eHealth usage in Europe: A multidimensional approach from an online survey of 13,000 European Union internet users. J Med Internet Res 2016 Jul 22;18(7):e188 [FREE Full text] [doi: 10.2196/jmir.5605] [Medline: 27450189]
27. Haluza D, Wernhart A. Does gender matter? Exploring perceptions regarding health technologies among employees and students at a medical university. Int J Med Inform 2019 Oct;130:103948. [doi: 10.1016/j.ijmedinf.2019.08.008] [Medline: 31442846]
28. Yamaguchi S, Waki K, Nannya Y, Nangaku M, Kadowaki T, Ohe K. Usage patterns of GlucoNote, a self-management smartphone app, based on ResearchKit for patients with type 2 diabetes and prediabetes. JMIR Mhealth Uhealth 2019 Apr 24;7(4):e13204 [FREE Full text] [doi: 10.2196/13204] [Medline: 31017586]
29. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: Insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014 Jul 16;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]

30. Ono H, ZavodnY M. Gender differences in information technology usage: A US-Japan Comparison. Sociol Perspec 2005 Mar 01;48(1):105-133. [doi: 10.1525/sop.2005.48.1.105]
31. Centis E, Trento M, Dei Cas A, Pontiroli AE, De Feo P, Bruno A, et al. Stage of change and motivation to healthy diet and habitual physical activity in type 2 diabetes. Acta Diabetol 2014 Aug;51(4):559-566. [doi: 10.1007/s00592-013-0551-1] [Medline: 24442514]
32. Huttunen-Lenz M, Hansen S, Christensen P, Meinert Larsen T, Sandø-Pedersen F, Drummen M, et al. PREVIEW study-influence of a behavior modification intervention (PREMIT) in over 2300 people with pre-diabetes: Intention, self-efficacy and outcome expectancies during the early phase of a lifestyle intervention. Psychol Res Behav Manag 2018;11:383-394 [FREE Full text] [doi: 10.2147/PRBM.S160355] [Medline: 30254498]
33. Prochaska JO, DiClemente CC, Norcross JC. In search of how people change. Applications to addictive behaviors. Am Psychol 1992 Sep;47(9):1102-1114. [doi: 10.1037//0003-066x.47.9.1102] [Medline: 1329589]
34. Kim C, Hwang A, Yoo J. The impact of a stage-matched intervention to promote exercise behavior in participants with type 2 diabetes. Int J Nurs Stud 2004 Nov;41(8):833-841. [doi: 10.1016/j.ijnurstu.2004.03.009] [Medline: 15476756]
35. Jalilian M, Moeini B, Hazavehei SM, Beigi AM, Sharifirad G, Nodeh FH. Physical activity stage-matched intervention: Promoting metabolic control in type 2 diabetes. J Educ Health Promot 2013;2:18 [FREE Full text] [doi: 10.4103/2277-9531.112688] [Medline: 24083268]
36. Friedberg JP, Rodriguez MA, Watsula ME, Lin I, Wylie-Rosett J, Allegrante JP, et al. Effectiveness of a tailored behavioral intervention to improve hypertension control: Primary outcomes of a randomized controlled trial. Hypertension 2015 Feb;65(2):440-446. [doi: 10.1161/HYPERTENSIONAHA.114.03483] [Medline: 25403606]
37. Romain AJ, Bortolon C, Gourlan M, Carayol M, Decker E, Lareyre O, et al. Matched or nonmatched interventions based on the transtheoretical model to promote physical activity. A meta-analysis of randomized controlled trials. J Sport Health Sci 2018 Jan;7(1):50-57 [FREE Full text] [doi: 10.1016/j.jshs.2016.10.007] [Medline: 30356472]
38. Wang Q, Egelandsdal B, Amdam GV, Almli VL, Oostindjer M. Diet and physical activity apps: Perceived effectiveness by app users. JMIR Mhealth Uhealth 2016 Apr 07;4(2):e33 [FREE Full text] [doi: 10.2196/mhealth.5114] [Medline: 27056639]
39. Fujita M, Sato Y, Nagashima K, Takahashi S, Hata A. Validity assessment of self-reported medication use by comparing to pharmacy insurance claims. BMJ Open 2015 Nov 09;5(11):e009490 [FREE Full text] [doi: 10.1136/bmjopen-2015-009490] [Medline: 26553839]

Abbreviations

eHealth: electronic health
EZR: Easy R
ICT: information and communication technology
OR: odds ratio
PHR: personal health record

Edited by G Eysenbach; submitted 18.03.19; peer-reviewed by A Pfammatter, M Lee; comments to author 09.10.19; revised version received 03.12.19; accepted 25.03.20; published 28.05.20.

Please cite as:
Hirano R, Yamaguchi S, Waki K, Kimura Y, Chin K, Nannya Y, Nangaku M, Kadowaki T, Ohe K
Willingness of Patients Prescribed Medications for Lifestyle-Related Diseases to Use Personal Health Records: Questionnaire Study
J Med Internet Res 2020;22(5):e13866
URL: https://www.jmir.org/2020/5/e13866
doi:10.2196/13866
PMID:32463368

Original Paper

A Blended Learning System to Improve Motivation, Mood State, and Satisfaction in Undergraduate Students: Randomized Controlled Trial

Mario Lozano-Lozano[1,2,3*], PhD; Carolina Fernández-Lao[1,2,3], PhD; Irene Cantarero-Villanueva[1,2,3*], PhD; Ignacio Noguerol[4], MSc; Francisco Álvarez-Salvago[1], PhD; Mayra Cruz-Fernández[1], MSc; Manuel Arroyo-Morales[1,2,3,5], MD, PhD; Noelia Galiano-Castillo[1,2,3], PhD

[1]Department of Physical Therapy, University of Granada, Granada, Spain

[2]Sport and Health University Research Institute, University of Granada, Granada, Spain

[3]Biohealth Research Institute in Granada, Granada, Spain

[4]Departamento de Lenguajes y Sistemas Informáticos e Ingeniería de Software, Universidad Politécnica de Madrid, Madrid, Spain

[5]Scientific Unit of Excellence on Exercise and Health, University of Granada, Granada, Spain

[*]these authors contributed equally

Corresponding Author:
Carolina Fernández-Lao, PhD
Department of Physical Therapy
University of Granada
Avda de la Ilustración, 60
Granada, 18016
Spain
Phone: 34 958248037
Email: carolinafl@ugr.es

Abstract

Background: Smartphone-based learning, or mobile learning (m-learning), has become a popular learning-and-teaching strategy in educational environments. Blended learning combines strategies such as m-learning with conventional learning to offer continuous training, anytime and anywhere, via innovative learning activities.

Objective: The main aim of this work was to examine the short-term (ie, 2-week) effects of a blended learning method using traditional materials plus a mobile app—the iPOT mobile learning app—on knowledge, motivation, mood state, and satisfaction among undergraduate students enrolled in a health science first-degree program.

Methods: The study was designed as a two-armed, prospective, single-blind, randomized controlled trial. Subjects who met the inclusion criteria were randomly assigned to either the intervention group (ie, blended learning involving traditional lectures plus m-learning via the use of the iPOT app) or the control group (ie, traditional on-site learning). For both groups, the educational program involved 13 lessons on basic health science. The iPOT app is a hybrid, multiplatform (ie, iOS and Android) smartphone app with an interactive teacher-student interface. Outcomes were measured via multiple-choice questions (ie, knowledge), the Instructional Materials Motivation Survey (ie, motivation), the Profile of Mood States scale (ie, mood state), and Likert-type questionnaires (ie, satisfaction and linguistic competence).

Results: A total of 99 students were enrolled, with 49 (49%) in the intervention group and 50 (51%) in the control group. No difference was seen between the two groups in terms of theoretical knowledge gain (P=.92). However, the intervention group subjects returned significantly higher scores than the control group subjects for all postintervention assessed items via the motivation questionnaire (all P<.001). Analysis of covariance (ANCOVA) revealed a significant difference in the confusion and bewilderment component in favor of the intervention group (P=.01), but only a trend toward significance in anger and hostility as well as total score. The intervention group subjects were more satisfied than the members of the control group with respect to five out of the six items evaluated: general satisfaction (P<.001), clarity of the instructions (P<.01), clarity with the use of the learning method (P<.001), enough time to complete the proposed exercises (P<.01), and improvement in the capacity to learn content (P<.001). Finally, the intervention group subjects who were frequent users of the app showed stronger motivation, as well as increased perception of greater gains in their English-language competence, than did infrequent users.

Conclusions: The blended learning method led to significant improvements in motivation, mood state, and satisfaction compared to traditional teaching, and elicited statements of subjective improvement in terms of competence in English.

Trial Registration: ClinicalTrials.gov NCT03335397; https://clinicaltrials.gov/ct2/show/NCT03335397

(J Med Internet Res 2020;22(5):e17101) doi:10.2196/17101

KEYWORDS

learning; teaching; students, health occupations; mobile apps; education

Introduction

Smartphone-based learning or mobile learning (m-learning), a term that highlights the type of device used, has become a popular learning-and-teaching strategy. It has been defined as the ability to access educational resources, tools, and materials anytime and from anywhere, using a mobile device, such as a smartphone [1]. With respect to health professionals and their education, Dunleavy et al describe m-learning as "any intervention using handheld, mobile devices connected through wireless connections to deliver educational content to pre and postregistration health professionals in order to extend the reach of learning and teaching beyond physical space and distance" [2]. M-learning thus provides a self-directed learning environment that affords continuous access to knowledge, information, and practice tools.

M-learning is being increasingly used by undergraduate and postgraduate students undertaking specialty or continuous training in the health sciences [2,3]. A recent review and meta-analysis showed that m-learning is equally or more effective than traditional learning in this field, although the authors also comment on the need for further research into its value [2]. Given the rapid evolution of mobile technologies in general, and of m-learning in particular, up-to-date evidence is required for the effectiveness of these strategies to be determined.

Motivation is an important factor in learning and performance [4]. Motivation levels can vary depending on learning-and-teaching style. A positive relationship has been reported between motivation and academic performance and learning, while a negative association is associated with dropping out from education [5]. Thus, motivation can be increased if the learning-and-teaching style is appropriate [6,7]. M-learning incorporates different activity options, which students can choose depending on their personal learning preferences or circumstances. This can encourage motivation, engagement, and learning success, positively influencing mood status [6,7].

New technologies may, however, also have a negative effect on learning. A recent study [8] highlights how students believed that new technologies in the classroom might affect their concentration and ability to learn. However, other investigations have found that students using electronic devices during a lecture returned results in learning that were similar to those who did not use these devices [9]. In other work, using electronic devices as learning tools while listening to a teacher was reported to not be distracting [10]. Clearly, the impact of the use of m-learning needs to be further investigated.

Blended learning that combines m-learning with conventional learning to seek the benefits of both [11,12] can also improve the educational experience [13], allowing conventionally taught material to be revisited whenever and wherever the student wishes [12,14]. Previous investigations into blended learning have returned positive results [15,16]. The main aim of this work was to examine the short-term effects of a blended learning method using traditional materials plus a mobile app on specific outcomes among students enrolled in a health science first-degree program. It was hypothesized that, compared to traditional learning alone, a 2-week blended learning program would improve knowledge uptake, student motivation, mood, and satisfaction and would improve competence in English.

Methods

Study Design and Sample Size

This study was a two-arm, prospective, single-blind, randomized controlled trial (ClinicalTrials.gov identifier: NCT03335397). The study subjects were 99 students in their first year of a health sciences degree at the Faculty of Health Sciences, University of Granada, Spain.

The sample size calculation was based on a previous study [17] and a demand of 90% power to detect a difference of 1.70 points in knowledge gain (see Main Outcomes section), while assuming a type 1 error (alpha) of 5% and a type 2 error (beta) of 10%, determined using G*Power software, version 3.1.9.3 (Heinrich-Heine-Universität Düsseldorf), for Mac OS X [18]. The required sample size was estimated at 40 subjects each for the intervention and control groups. Contemplating a dropout rate of 25%, as it is common for health science students to migrate to medicine in the first semester of their first academic year, 50 subjects were enrolled per group.

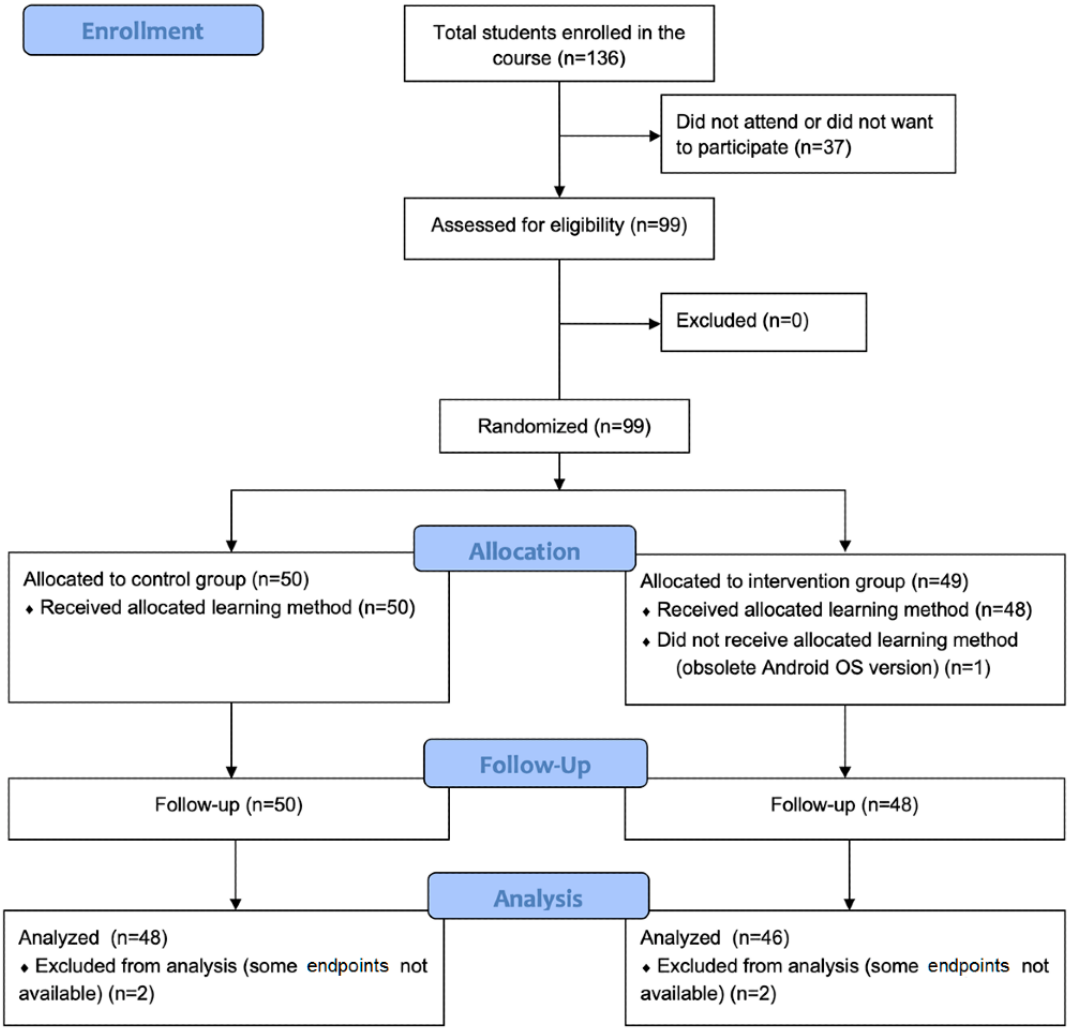
Recruitment

Subjects were recruited via a talk given on the first day of the course unit entitled Basic Health Science, a unit lasting one semester. It was emphasized that no assessment made during the trial would have any effect on final grades and that participation was totally voluntary. To be included, the subjects had to (1) possess basic skills in handling mobile apps, (2) have a smartphone running either Android operating system (OS) or iOS software, (3) install the iPOT mobile learning app, (4) be enrolled in the above course unit, and (5) have a basic knowledge of English, accredited or not. In addition, all subjects had to provide informed consent to be included. Students repeating the unit were excluded.

Randomization

Subjects who met the inclusion criteria were randomly assigned to either the intervention group (ie, blended learning, involving traditional learning plus m-learning via the use of the iPOT app) or the control group (ie, traditional on-site learning). Figure 1 shows a flowchart of the recruitment and randomization process. Subjects were allocated to these groups via the computer generation of a random number list and the subsequent production of a subject allocation sequence using SPSS statistics software, version 22.0 (IBM Corp). The subjects then met with the evaluator who would conduct the forthcoming assessments; this evaluator was blinded to all group assignments.

Figure 1. Flow diagram of the recruitment and randomization process. OS: operation system.



Experimental Procedure

This study was conducted over 2 weeks in the first semester of the 2017-2018 academic year. For both groups, the educational program involved 13 lessons on basic health science. The subjects of both groups were progressively introduced to conventional learning materials (ie, books, PowerPoint presentations, and journals available in the university library). The subjects in the intervention group, however, also received the iPOT app to reinforce the educational program; it provided no additional material. Both the intervention and control subjects were free to reinforce the taught information using any of the traditional sources available. All subjects in the intervention group received a QR (Quick Response) code to provide them access to the Apple App Store [19] or the Google Play store [20] as required. Subjects then installed the iPOT app. A lecturer involved in the project was present at this time to help solve any technical problems that might arise. All the intervention group subjects were provided a personal password linked to their institutional email account that allowed them to access the app, but were also assured each person's log-in information remained confidential and nonexchangeable with subjects in the control group; the system itself was enabled to detect irregular connections. However, given that students in the control group might borrow the devices of the intervention group members in order to try the app out, all subjects were requested to respect the protocol and told they would have unlimited access to the app at the end of the study. The iPOT app included a permanently available video tutorial (see Multimedia Appendix 1), accessible through the *Help* tab, to assist users in familiarizing themselves with the app and its operation. In addition, a WhatsApp group was established for the 2 weeks of the experimental period in order to resolve any incidents that might arise. Two smartphones were available to

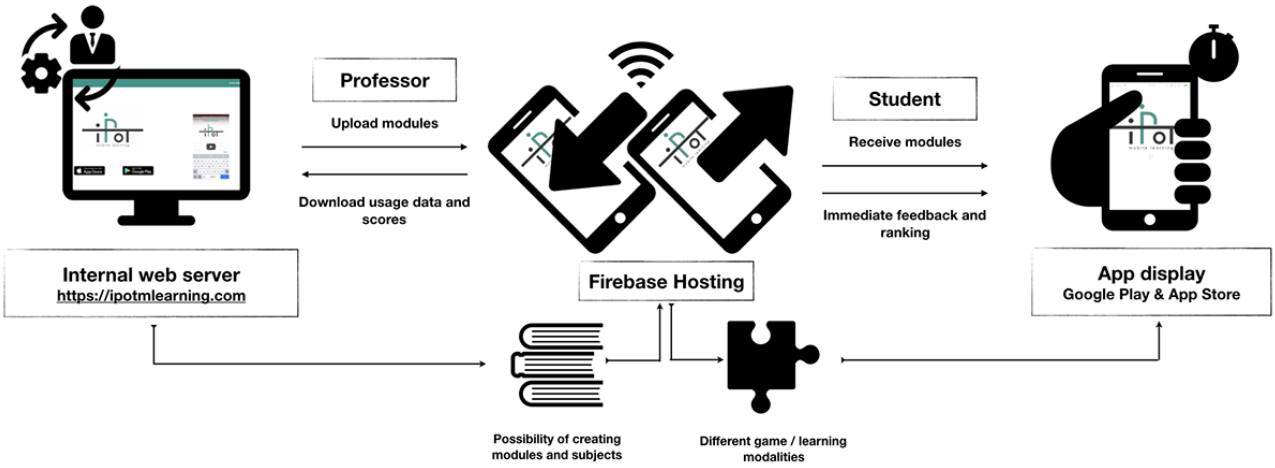
loan to students who wanted to participate but who had devices with compatibility problems.

M-learning lessons were enabled in the app over the last 2 weeks of the teaching period, allowing the intervention group subjects to review the information previously taught. Students were instructed to review each lesson via the six learning modalities available—quizzes (multiple-choice questions), alphabet soup (word search), hangman, hieroglyphics, puzzles (word sorting), and word-coupling activities—and were encouraged to use them; differences in use time were accounted for (see The iPOT Mobile Learning App: Internal Design and Technical Specifications section). It is important to highlight that during these 2 weeks, the control group subjects were encouraged to review the taught material via the above-mentioned books, PowerPoint presentations, and journals.

The iPOT Mobile Learning App: Internal Design and Technical Specifications

The iPOT app was designed by a research group composed of lecturers in health science and an internship student of computer engineering, all at the University of Granada. The app is a hybrid, multiplatform (iOS and Android), smartphone app with an interactive teacher-student interface (see Figure 2). Its architecture involves Firebase, a Google cloud-based development platform that handles user authentication, file storage, and database management. The app interface gives users access to different game modes and allows them to complete different activities and find platform-related information. The app also has an administration console, developed using Angular and Angular Material development tools (Google), that gives administrators the possibility of managing users, all app modules, game-related data, and graphic resources by adding an abstraction layer between the administrator and Firebase. For the deployment of the administration console, it was decided to use Firebase Hosting, a static, rapid hosting service that includes a Secure Sockets Layer (SSL) by default, ensuring safe access.

Figure 2. Top-level view of the iPOT app system.



The user interface shows the user profile, allowing access to personal data and providing the opportunity to change one's nickname; ranking (ie, position in terms of game scores; designed to help increase motivation via a sense of competition); and a start button to allow users to begin playing. As an incentive, each time a correct answer was given in any of the app's modalities, points were added to the user's score. Each user's scores, as well as the number of modules completed, were recorded via the system.

Outcome Measures

Knowledge and mood state were assessed for both the intervention and control groups before and after the 2-week experimental period. Motivation, satisfaction, and English-language competence—the latter only measured in the intervention group—were assessed after this period.

Knowledge

Theoretical knowledge was evaluated at baseline (ie, the beginning of the 2-week study period) and at the end of the study period using 20 multiple-choice questions [21]. This tool has been used in other educational studies [22,23]. Scores were recorded on an 11-point scale, ranging from 0 to 10, with higher values representing better outcomes. This test was developed by two external lecturers who were not involved in the study.

Motivation

Motivation was assessed using the Instructional Materials Motivation Survey (IMMS) [24]. The terminology used by the tool was adapted to the learning methods employed in this study. The survey can be used to estimate the motivation and attitudes of students in situations of self-directed learning (ie, printed, virtual, or online courses). It contains 36 questions in four domains: attention, relevance, confidence, and satisfaction. Each question is scored on a 5-point Likert scale; therefore, total scores range from 36 to 180. Higher scores denote higher levels of learning motivation. The tool has been shown to be highly reliable (intraclass correlation coefficient [ICC]>.96) [25] and has been previously used in Spanish populations [26,27].

Mood State

Mood is a short-term state of feeling; it may fluctuate within minutes or over days. Unlike emotions, moods are more transient, often unrelated to external events, and are of varying intensity [28]. Subject mood state was assessed via the Profile of Mood States (POMS) questionnaire [29], validated for use in Spain [30]. This self-reported survey contains 65 questions, each scored on an increasing 5-point Likert scale, ranging from 0 to 4, that together measure six mood components: (1) tension and anxiety (nine items that refer to an increase in musculoskeletal tension, somatic tension, and observable psychomotor manifestations); (2) depression and melancholy (15 items that reflect a depressed state along with a feeling of sadness or guilt); (3) anger and hostility (12 items that represent feelings of anger and antipathy toward others); (4) vigor and activity (eight items that allude to a state of euphoria or high energy); (5) fatigue and inertia (seven items that refer to a mood of depression or a low level of exercise); and (6) confusion and bewilderment (seven items referring to disorientation or multiplicity of thought). The total average score of all the components (ie, the Total Mood Disturbance [TMD] score) was recorded. There are no cutoff points in any domain. Subjects were asked to fill in the questionnaire indicating how they felt lately. Lower values indicate a better general mood state [31]. The test is reliable; the ICC is close to .90 for all six components [29].

Satisfaction

Subjects' satisfaction with the learning methods to which they were exposed was evaluated via a questionnaire based on that reported by Brewer et al [32], with six items that assess satisfaction variables: general satisfaction, the clarity of instructions, whether the final assessment reflected the course syllabus, clarity with the use of the learning method, whether there was enough time to complete the proposed exercises, and improvement in the capacity to learn content. Each item was scored on an increasing 5-point Likert scale, ranging from 1 to 5. Higher values represent better outcomes. Subjects were also asked to report the number of hours per day, as well as the number of days per week, that they used the app and which mode of the app they preferred.

Linguistic Competence

The subjective improvement in English-language competence (ie, a general feeling of improvement and of improvements in written comprehension, written expression, vocabulary, and global perception) was measured at the end of the study via an ad hoc questionnaire, scoring answers on an increasing 5-point Likert scale, ranging from 1 to 5. Higher values represent better outcomes. To determine the subjects' prior English-language competence, they were asked about any official English certificates they possessed before starting the study.

Statistical Analysis

The normal distribution of data was verified using the Kolmogorov-Smirnov test. Results are presented as mean (SD) values. The Student *t* test and Mann-Whitney U test or the chi-square test and Fisher exact test were used as required to examine differences in continuous and categorical variables, respectively, between the two groups (eg, sociodemographic characteristics, knowledge, and mood state) and in postintervention differences (eg, motivation, satisfaction, and English-language competence).

Differences between the groups in terms of the preintervention-to-postintervention change in knowledge and mood state were analyzed by repeated-measures analysis of covariance (ANCOVA), adjusting for the effects of those variables showing significant differences between groups at baseline (ie, the covariate *POMS-vigor and activity*). Missing values were few (<5% of the total number); in such cases, these can be considered *missing at random* and inconsequential [33]. For this reason, a list-wise deletion method was chosen; no multiple imputation was necessary. Significance was set at $P<.05$. All calculations were performed using SPSS statistics software, version 22.0 (IBM Corp), and Stata statistical software, release 14 (StataCorp).

Ethical Approval

The study was performed in accordance with the ethical standards of the appropriate national and institutional research committees, and adhering to the Declaration of Helsinki [34]. Ethical approval was granted as required by Spanish Law 223/2004, of February 6, establishing subject confidentiality under the terms of Spanish Law 3/2018, of December 5 [35]. The study was also approved by the Quality, Innovation and Planning Unit of the University of Granada, Spain (Plan FIDO UGR 2016-2018). For ethical reasons, once the experimental phase of the project was complete, the app was made available to all subjects who took part.

Results

User Statistics

A total of 99 students were enrolled, with 49 (49%) in the intervention group and 50 (51%) in the control group. The mean age of the students was 19.76 years (SD 2.74) in the intervention group versus 20.00 years (SD 3.98) in the control group ($P=.97$). A total of 38 of the 49 subjects (78%) in the intervention group were women compared to 35 out of 50 subjects (70%) in the control group ($P=.39$). A total of 37 of the 49 subjects (76%) in the intervention group used mobile phones with the Android OS compared to 31 out of 50 subjects (62%) in the control group ($P=.14$). No significant differences were seen between the members of the two groups in terms of their baseline levels of English ($P=.69$) or baseline knowledge of the material to be taught ($P=.43$) (see Table 1). English levels were determined according to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment [36]. The intervention group lost one subject (1/49, 2%) to follow-up—a dropout due to the subject's phone running an old OS that did not support the iPOT app—leaving a final initial sample size of 48. Another 2 subjects were lost from both groups due to incompletion of the final assessments (see Figure 1).

Table 1. Sociodemographic characteristics of the study subjects at baseline: the beginning of the 2-week study period.

Characteristic	Control group (n=50)	Intervention group (n=49)
Age (years), mean (SD)	20.00 (3.98)	19.76 (2.74)
Gender, n (%)		
Male	15 (30)	11 (22)
Female	35 (70)	38 (78)
English level, n (%)[a]		
Not certified	25 (50)	22 (45)
A2 (elementary)	4 (8)	2 (4)
B1 (low intermediate)	14 (28)	15 (31)
B2 (high intermediate)	5 (10)	9 (18)
C1 (advanced)	2 (4)	1 (2)
Mobile system, n (%)		
iOS	19 (38)	12 (24)
Android operating system (OS)	31 (62)	37 (76)
Baseline knowledge of subjects (score[b]), mean (SD)	4.38 (1.21)	4.57 (1.22)

[a]English levels were determined according to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment [36]. A2: Basic User, Waystage; B1: Independent User, Threshold; B2: Independent User, Vantage; C1: Proficient User, Effective Operational Proficiency.

[b]Scores ranged from 0 (no knowledge) to 10 (highest level of knowledge).

Main Outcomes

At the end of the experimental period, no significant difference was seen between the two groups in terms of theoretical knowledge gain (intervention group mean score 0.51 [SD 1.19] vs control group mean score 0.49 [SD 1.10]; F_1=0.008, P=.92) (see Table 2). A significant difference was also not seen in the end-of-study total scores (mean score 5.16 [SD 1.30] vs mean score 4.90 [SD 1.24], respectively; t_{92}=-1.019, P=.31). The covariate *POMS-vigor and activity* did not influence the results.

Table 3 shows that the intervention group subjects returned significantly higher scores than the control group subjects for all items assessed postintervention by the IMMS (IMMS_attention: t_{92}=–12.223; IMMS_relevance: t_{92}=–8.315; IMMS_confidence: t_{92}=–7.731; and IMMS_satisfaction: t_{92}=–10.631; P<.001 for all). Their IMMS total motivation scores were also significantly different (mean 136.02 [SD 19.25] vs mean 84.94 [SD 25.37], respectively; t_{92}=–10.963, P<.001).

Table 4 reflects the mood state of the subjects in each group before and after the intervention. ANCOVA revealed a significant difference between the change in confusion and bewilderment for the control and intervention group subjects (mean score 3.67 [SD 8.66] vs mean score –0.50 [SD 6.62], respectively; F_1=6.826, P=.01). A trend toward significance was also seen between the groups in terms of the change in anger and hostility (control group mean score 2.21 [SD 7.97] vs intervention group mean score –1.07 [SD 8.45]; F_1=3.735, P=.05) and total score (control group mean score –1154.17 [SD 3652.10] vs intervention group mean score –17.39 [SD 2767.05]; F_1=2.875, P=.09). No significant differences between groups were seen in terms of the change in tension and anxiety (F_1=1.964, P=.16), depression and melancholy (F_1=0.983, P=.32), vigor and activity (F_1=0.181, P=.67), and fatigue and inertia (F_1=0.233, P=.63). Adjustment for the covariate POMS-vigor and activity in ANCOVA revealed this variable to influence the change in confusion and bewilderment (F_1=7.889, P=.006), anger and hostility (F_1=4.493, P=.03), and total score (F_1=4.603, P=.03). The covariate did not influence the rest of results.

Table 2. Effect of the different learning methods on knowledge gain.

Answers at study time points	Control group (n=48), mean (SD)	Intervention group (n=46), mean (SD)	*P* value[a]
Correct answers			.92
Baseline[b]	8.81 (2.43)	9.30 (2.37)	
Postintervention	9.79 (2.48)	10.33 (2.61)	
Difference	0.98 (2.20)	1.02 (2.37)	
Wrong answers			.92
Baseline	11.19 (2.43)	10.70 (2.38)	
Postintervention	10.21 (2.48)	9.67 (2.61)	
Difference	–0.98 (2.20)	–1.02 (2.37)	
Total score			.92
Baseline	4.41 (1.21)	4.65 (1.19)	
Postintervention	4.90 (1.24)	5.16 (1.30)	
Difference	0.49 (1.10)	0.51 (1.19)	

[a]Repeated-measures analysis of covariance (ANCOVA) was used to examine the differences between groups.

[b]*Baseline* refers to the beginning of the 2-week study period.

Table 3. Learning motivation in the intervention and control groups.

Learning motivation variable[a,b]	Control group (n=48), mean (SD)	Intervention group (n=46), mean (SD)	*P* value[c]
Attention (12-60)	26.06 (9.64)	47.59 (7.20)	<.001
Relevance (9-45)	24.06 (6.58)	34.17 (5.08)	<.001
Confidence (9-45)	22.04 (6.96)	31.83 (5.13)	<.001
Satisfaction (6-30)	12.77 (4.55)	22.44 (4.25)	<.001
Total IMMS[a] score (36-180)	84.94 (25.37)	136.02 (19.25)	<.001

[a]Learning motivation was measured using the Instructional Materials Motivation Survey (IMMS).

[b]Each category is presented with the possible range of its score in parentheses.

[c]The Student *t* test was used to examine the differences between groups. Significance was set at *P*<.05.

Figure 3 shows the results of the global satisfaction survey with the learning method used. The intervention group subjects were more satisfied than the members of the control group in terms of all items evaluated: general satisfaction with the learning method (*Strongly agreed* and *Agreed* in the intervention group were 30% and 65%, respectively, vs 2% and 10% in the control group; *P*<.001), clarity of the instructions (*Strongly agreed* and *Agreed* in the intervention group were 57% and 35%, respectively, vs 27% and 31% in the control group; *P*<.01), clarity with the use of the learning method (*Strongly agreed* and *Agreed* in the intervention group were 30% and 54%, respectively, vs 2% and 29% in the control group; *P*<.001), enough time to complete the proposed exercises (*Strongly agreed* and *Agreed* in the intervention group were 9% and 17%, respectively, vs 8% and 4% in the control group; *P*<.01), and improvement in the capacity to learn content (*Strongly agreed* and *Agreed* in the intervention group were 11% and 57%, respectively, vs 0% and 17% in the control group; *P*<.001). No difference was seen between the groups with respect to satisfaction regarding how the final assessment reflected the course syllabus (*Strongly agreed* and *Agreed* in the intervention group were 11% and 43%, respectively, vs 6% and 31% for the control group; *P*=.06).

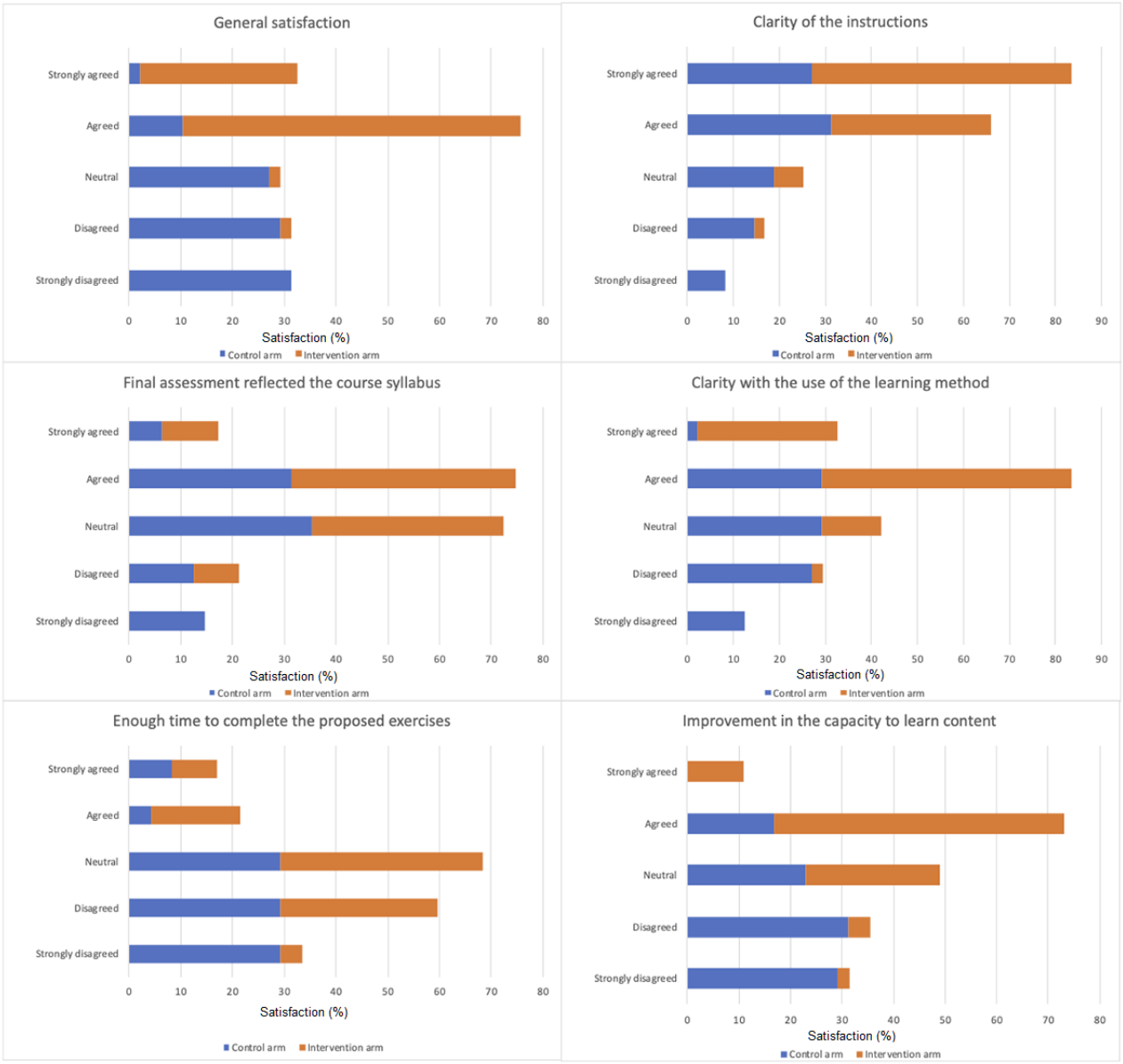
Table 4. Effect of the different learning methods on the Profile of Mood States (POMS) total and subscale scores.

Mood state	Control group (n=48), mean (SD)	Intervention group (n=46), mean (SD)	*P* value[a]
Tension and anxiety			.16
Baseline[b]	45.58 (8.43)	46.78 (9.70)	
Postintervention	46.15 (9.45)	45.00 (10.09)	
Difference	0.56 (8.88)	–1.78 (7.21)	
Depression and melancholy			.32
Baseline	47.38 (5.55)	47.39 (6.82)	
Postintervention	49.21 (7.64)	47.93 (8.64)	
Difference	1.83 (6.55)	0.54 (6.04)	
Anger and hostility			.05
Baseline	51.77 (9.52)	52.67 (8.95)	
Postintervention	53.98 (11.79)	51.61 (10.24)	
Difference	2.21 (7.97)	–1.07 (8.45)	
Vigor and activity			.67
Baseline	54.25 (7.30)	57.43 (6.05)	
Postintervention	51.96 (7.55)	54.67 (6.77)	
Difference	–2.29 (5.80)	–2.76 (4.82)	
Fatigue and inertia			.63
Baseline	47.75 (7.02)	46.59 (6.79)	
Postintervention	48.73 (9.49)	46.80 (7.45)	
Difference	0.98 (9.24)	0.22 (5.51)	
Confusion and bewilderment			.01
Baseline	38.54 (7.38)	39.65 (8.02)	
Postintervention	42.21 (8.43)	39.15 (7.64)	
Difference	3.67 (8.66)	–0.50 (6.62)	
Total score			.09
Baseline	–17677.08 (2899.91)	–17565.22 (3453.13)	
Postintervention	–18831.25 (4052.59)	–17582.61 (4088.16)	
Difference	–1154.17 (3652.10)	–17.39 (2767.05)	

[a]Repeated-measures analysis of covariance (ANCOVA) was used to examine the differences between groups. Significance was set at $P<.05$; a trend toward significance was defined as $.05 \leq P < .10$.

[b]*Baseline* refers to the beginning of the 2-week study period.

Figure 3. Satisfaction among the intervention and control arms.



The intervention group subjects used the app an average of 3.85 (SD 1.51) days per week. The *quiz* option was the most popular, used by 83% (38/46) of the subjects, followed by *word coupling* (29/46, 63%) and *alphabet soup* (20/46, 43%). In further analysis, the median number of days spent using the app per week (median 4 days) was used to categorize students as infrequent (<4 days/week) and frequent users (≥4 days/week). Motivation among the frequent users was significantly higher than among the infrequent users (IMMS_attention: frequent, mean score 50.04 [SD 5.17] vs infrequent, mean score 44.40 [SD 8.29], t_{92}=–2.669, *P*=.01; IMMS_confidence: frequent, mean score 33.77 [SD 4.87] vs infrequent, mean score 29.30 [SD 4.38], t_{92}=–3.222, *P*<.01; IMMS_satisfaction: frequent, mean score 24.00 [SD 3.03] vs infrequent, mean score 20.40 [SD 4.78], t_{92}=–3.114, *P*<.01; and IMMS_total: frequent, mean score 142.85 [SD 15.52] vs infrequent, mean score 127.15 [SD 20.34], t_{92}=–2.971, *P*<.01). No significant difference was detected, however, between the users in the subgroups in terms of IMMS_relevance (frequent, mean score 35.04 [SD 4.24] vs infrequent, mean score 33.05 [SD 5.92], t_{92}=–1.328, *P*=.19). The frequent users also perceived their general linguistic competence in English to have significantly improved through the use of the app (*P*=.03), although this was not reflected when examining specific domains: written expression (*P*=.57), written comprehension (*P*=.18), and vocabulary (*P*=.11).

Discussion

Principal Findings

The blended learning method led to significant improvements in motivation, mood state, and satisfaction compared to traditional teaching. Moreover, frequent users of the app showed stronger motivation and perceived greater gains in their English-language competence than did infrequent users. However, the results suggest the intervention strategy provided no benefit in terms of knowledge absorbed. The latter finding is consistent with the results of two previous randomized controlled studies evaluating online learning methods [22,23].

In contrast, Chuang et al, whose work had a similar design to this investigation (ie, blended learning vs regular lecturing) and was of the same duration (ie, 2 weeks), reported significant differences in favor of blended learning [37]. It has been suggested that the use of audiovisual material, as employed in the latter study, might generate better learning outcomes [38].

Comparison With Prior Work

In this work, motivation was measured by the validated IMMS tool. Higher scores were recorded for all motivation dimensions in the intervention group. Keller defines motivation as an innate characteristic of students, but also indicates that it can be influenced by external factors such as the instructional method used [39]. The innovative design for reviewing classes described in this work may be responsible for the increased motivation among the intervention group subjects. The use of augmented reality among middle school students [26], and of 3D computer environments among undergraduate students [40], have returned similar improvements in motivation. A recent study [41] reports that a mobile augmented reality (mAR) app boosted motivation in 35 students at four vocational education and training institutes over a period of 20 days. Real-time feedback, the degree of success achieved (ie, ranking), time on task (ie, frequent vs infrequent users), and learning outcomes (ie, theoretical tests) were some of the main predictors of improved learning in this latter study. The greatest difference between the subjects in the groups of this study was seen with respect to IMMS_attention (ie, the interest and curiosity of students in the learning process), which was significantly greater among the intervention subjects. The app's attractive interface, enjoyable quizzes, and user-friendliness may be partially responsible for this result.

The intervention group subjects appeared to be more involved in the learning process; they returned a significant change in two of the six mood state components assessed, which was reflected in the total score. Moreover, anger and hostility as well as confusion and bewilderment improved in the intervention group subjects, while these became worse (ie, increased scores) among those in the control group. The use of the app might thus have helped improve negative mood states. Pekrun et al [42] stated that academic achievement is a complex interaction between emotions, engagement, and performance, and this influences student self-appraisal. Flexible and creative learning strategies facilitate positive academic emotions, while more rigid strategies may spur negative emotions [42]. The “flow” concept proposes that an optimal educational experience may exist when students feel that a task is meaningful and, while perhaps challenging, that they are up to the challenge [43]. Other authors [17,44] who tested mAR interventions (ie, 30 minutes and 45 minutes in length), using a variation of the POMS questionnaire to measure change in mood state, only showed intrasubject changes.

Finally, satisfaction was greater among the intervention group subjects. It seems clear that the role of m-learning was perceived as an appropriate complement to traditional lecturing. With the development of advanced technologies, smartphones can be an effective learning tool for students. The possibility of reviewing classes whenever and wherever one likes, with tailored feedback (ie, scores for the games and quizzes), may have promoted the transfer of knowledge into their short-term and even long-term memories [45].

Limitations

This study suffers from the limitation that it lasted only 2 weeks. This short time frame was due to external factors such as a funder-imposed deadline. Further, the iPOT app was created for use as a post-teaching period review tool, not as a study tool in itself. No difference was detected between the amount of knowledge absorbed by the members of the two groups. However, if, as the results show, the use of the app motivates students, a significant difference in what is learned might be expected after a longer period of use. Further work should investigate this.

Conclusions

Students who received instruction via blended learning involving m-learning showed greater motivation, a better mood state, and greater satisfaction than those who received traditional lectures alone, although no difference was seen in terms of the amount of knowledge absorbed. Longer studies are needed to determine whether the improvements in these factors persist and whether they eventually translate into more knowledge being absorbed.

Acknowledgments

We are grateful to Adrian Burton for assistance with the English language. The authors are also grateful to the students who participated in the study. This work was funded by a grant from the Educational Innovation Unit of the University of Granada, Spain (PID 16-54). Additional funding was provided by the University of Granada, Plan Propio de Investigación 2016, Excellence actions: Units of Excellence; Unit of Excellence on Exercise and Health (UCEES).

Conflicts of Interest

None declared.

Multimedia Appendix 1
iPOT video tutorial.
[MP4 File (MP4 Video), 14574 KB - jmir_v22i5e17101_app1.MP4]

Multimedia Appendix 2

CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 387 KB - jmir_v22i5e17101_app2.pdf]

References

1. GSMA. mLearning: A Platform for Educational Opportunities at the Base of the Pyramid. London, UK: GSMA Development Fund; 2010 Nov. URL: https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2012/04/mlearningaplatformforeducationalopportunitiesatthebaseofthepyramid.pdf [accessed 2020-04-29]
2. Dunleavy G, Nikolaou CK, Nifakos S, Atun R, Law GC, Tudor Car L. Mobile digital education for health professions: Systematic review and meta-analysis by the digital health education collaboration. J Med Internet Res 2019 Feb 12;21(2):e12937 [FREE Full text] [doi: 10.2196/12937] [Medline: 30747711]
3. Guo P, Watts K, Wharrad H. An integrative review of the impact of mobile technologies used by healthcare professionals to support education and practice. Nurs Open 2016 Apr;3(2):66-78 [FREE Full text] [doi: 10.1002/nop2.37] [Medline: 27708817]
4. Kusurkar RA, Ten Cate TJ, van Asperen M, Croiset G. Motivation as an independent and a dependent variable in medical education: A review of the literature. Med Teach 2011 Apr 25;33(5):e242-e262. [doi: 10.3109/0142159x.2011.558539]
5. Wouters A, Croiset G, Galindo-Garre F, Kusurkar RA. Motivation of medical students: Selection by motivation or motivation by selection. BMC Med Educ 2016 Jan 29;16:37 [FREE Full text] [doi: 10.1186/s12909-016-0560-1] [Medline: 26825381]
6. Kilis S. Impacts of mobile learning in motivation, engagement and achievement of learners: Review of literature. Gaziantep Univ J Soc Sci 2013;12(2):375-383 ISSN: 1303-0094 [FREE Full text]
7. Orsini C, Binnie VI, Wilson SL. Determinants and outcomes of motivation in health professions education: A systematic review based on self-determination theory. J Educ Eval Health Prof 2016;13:19 [FREE Full text] [doi: 10.3352/jeehp.2016.13.19] [Medline: 27134006]
8. Attia N, Baig L, Marzouk YI, Khan A. The potential effect of technology and distractions on undergraduate students' concentration. Pak J Med Sci 2017;33(4):860-865 [FREE Full text] [doi: 10.12669/pjms.334.12560] [Medline: 29067054]
9. Nalliah R, Allareddy V. Students distracted by electronic devices perform at the same level as those who are focused on the lecture. PeerJ 2014;2:e572 [FREE Full text] [doi: 10.7717/peerj.572] [Medline: 25279260]
10. Pew Research Center. Millennials: A Portrait of Generation Next. Confident. Connected. Open to Change. Washington, DC: Pew Research Center; 2010 Feb. URL: https://www.pewresearch.org/wp-content/uploads/sites/3/2010/10/millennials-confident-connected-open-to-change.pdf [accessed 2020-04-29]
11. Delialioglu O, Yildirim Z. Design and development of a technology enhanced hybrid instruction based on MOLTA model: Its effectiveness in comparison to traditional instruction. Comput Educ 2008 Aug;51(1):474-483. [doi: 10.1016/j.compedu.2007.06.006]
12. Feist M, Ciccarelli M, McFerron BA, Molleston JP. Methods and effects of a case-based pediatric gastroenterology online curriculum. J Pediatr Gastroenterol Nutr 2013;56(2):161-165. [doi: 10.1097/mpg.0b013e31825677d7]
13. Graham CR. Blended learning systems: Definition, current trends, and future directions. In: Bonk CJ, Graham CR, editors. The Handbook Of Blended Learning: Global Perspectives, Local Designs. San Francisco, CA: Pfeiffer Publishing; 2005:3-21.
14. Yapici İ, Akbayin H. The effect of blended learning model on high school students' biology achievement and on their attitudes towards the internet. Turk Online J Educ Technol 2012;11(2):228-237 [FREE Full text]
15. Liu Q, Peng W, Zhang F, Hu R, Li Y, Yan W. The effectiveness of blended learning in health professions: Systematic review and meta-analysis. J Med Internet Res 2016 Jan 04;18(1):e2 [FREE Full text] [doi: 10.2196/jmir.4807] [Medline: 26729058]
16. Alamro AS, Schofield S. Supporting traditional PBL with online discussion forums: A study from Qassim Medical School. Med Teach 2012 Mar 12;34(sup1):S20-S24. [doi: 10.3109/0142159x.2012.656751]
17. Albrecht U, Folta-Schoofs K, Behrends M, von Jan U. Effects of mobile augmented reality learning compared to textbook learning on medical students: Randomized controlled pilot study. J Med Internet Res 2013 Aug 20;15(8):e182 [FREE Full text] [doi: 10.2196/jmir.2497] [Medline: 23963306]
18. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/BF03193146]
19. Apple App Store. iPOT URL: https://apps.apple.com/es/app/ipot/id1291766873 [accessed 2020-05-15]
20. Google Play. iPOT URL: https://play.google.com/store/apps/details?id=com.ipot.ipotMobileLearning&hl=es [accessed 2020-05-08]
21. Considine J, Botti M, Thomas S. Design, format, validity and reliability of multiple choice questions for use in nursing research and education. Collegian 2005 Jan;12(1):19-24. [doi: 10.1016/s1322-7696(08)60478-3] [Medline: 16619900]
22. Fernández-Lao C, Cantarero-Villanueva I, Galiano-Castillo N, Caro-Morán E, Díaz-Rodríguez L, Arroyo-Morales M. The effectiveness of a mobile application for the development of palpation and ultrasound imaging skills to supplement the traditional learning of physiotherapy students. BMC Med Educ 2016 Oct 19;16(1):274 [FREE Full text] [doi: 10.1186/s12909-016-0775-1] [Medline: 27756288]
23. Arroyo-Morales M, Cantarero-Villanueva I, Fernández-Lao C, Guirao-Piñeyro M, Castro-Martín E, Díaz-Rodríguez L. A blended learning approach to palpation and ultrasound imaging skills through supplementation of traditional classroom

teaching with an e-learning package. Man Ther 2012 Oct;17(5):474-478. [doi: 10.1016/j.math.2012.04.002] [Medline: 22579034]
24. Keller JM. Strategies for stimulating the motivation to learn. Nonprofit Manag Leadersh 1987 Oct;26(8):1-7. [doi: 10.1002/pfi.4160260802]
25. Keller JM. Motivational Design for Learning and Performance: The ARCS Model Approach. New York, NY: Springer Science+Business Media; 2010.
26. Di Serio Á, Ibáñez MB, Delgado Kloos C. Impact of an augmented reality system on students' motivation for a visual art course. Comput Educ 2013 Oct;68:586-596. [doi: 10.1016/j.compedu.2012.03.002]
27. Castaño-Garrido C, Maiz-Olazabalaga I, Garay-Ruiz U. Design, motivation and performance in a cooperative MOOC course. Comunicar 2015 Jan;22(44):19-26 [FREE Full text] [doi: 10.3916/c44-2015-02]
28. Berger BG, Motl RW. Exercise and mood: A selective review and synthesis of research employing the profile of mood states. J Appl Sport Psychol 2000 Mar;12(1):69-92. [doi: 10.1080/10413200008404214]
29. McNair DM, Lorr M, Droppleman LF. Manual for the Profile of Mood States. 3rd edition. San Diego, CA: Educational and Industrial Testing Service (EdITS); 1992.
30. Balaguer SI, Fuentes I, Meliá JL, García-Merita ML, Pérez Recio G. El perfil de los estados de ánimo (POMS): Baremo para estudiantes Valencianos y su aplicación en el contexto deportivo. Rev Psicol Deporte 1993;4(2):39-52 [FREE Full text]
31. Lane AM, Jackson A, Terry PC. Preferred modality influences on exercise-induced mood changes. J Sports Sci Med 2005 Jun 01;4(2):195-200 [FREE Full text] [Medline: 24431976]
32. Brewer ZE, Ogden WD, Fann JI, Burdon TA, Sheikh AY. Creation and global deployment of a mobile, application-based cognitive simulator for cardiac surgical procedures. Semin Thorac Cardiovasc Surg 2016;28(1):1-9. [doi: 10.1053/j.semtcvs.2016.02.006] [Medline: 27568126]
33. Schafer JL. Multiple imputation: A primer. Stat Methods Med Res 1999 Mar;8(1):3-15. [doi: 10.1177/096228029900800102] [Medline: 10347857]
34. World Medical Association. World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. JAMA 2013 Nov 27;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]
35. Spanish Government. Agencia Estatal Boletín Oficial del Estado. 2018 Dec 06. Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales URL: https://www.boe.es/eli/es/lo/2018/12/05/3 [accessed 2020-04-29]
36. Council of Europe. Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Cambridge, UK: Cambridge University Press; 2001. URL: https://rm.coe.int/1680459f97 [accessed 2020-05-08]
37. Chuang Y, Lai F, Chang C, Wan H. Effects of a skill demonstration video delivered by smartphone on facilitating nursing students' skill competencies and self-confidence: A randomized controlled trial study. Nurse Educ Today 2018 Jul;66:63-68. [doi: 10.1016/j.nedt.2018.03.027] [Medline: 29677580]
38. Mayer RE. Designing multimedia instruction in anatomy: An evidence-based approach. Clin Anat 2020 Jan;33(1):2-11. [doi: 10.1002/ca.23265] [Medline: 30144175]
39. Keller JM. Development and use of the ARCS model of instructional design. J Instr Dev 1987 Sep;10(3):2-10. [doi: 10.1007/bf02905780]
40. Hu A, Shewokis PA, Ting K, Fung K. Motivation in computer-assisted instruction. Laryngoscope 2016 Aug;126 Suppl 6:S5-S13. [doi: 10.1002/lary.26040] [Medline: 27307270]
41. Bacca J, Baldiris S, Fabregat R, Kinshuk. Insights into the factors influencing student motivation in augmented reality learning experiences in vocational education and training. Front Psychol 2018;9:1486 [FREE Full text] [doi: 10.3389/fpsyg.2018.01486] [Medline: 30186196]
42. Pekrun R, Goetz T, Titz W, Perry RP. Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. Educ Psychol 2002 Jun;37(2):91-105. [doi: 10.1207/s15326985ep3702_4]
43. Csikszentmihalyi M. Flow: The Psychology of Optimal Experience. New York, NY: HarperCollins Publishers; 2008.
44. Noll C, von Jan U, Raap U, Albrecht U. Mobile augmented reality as a feature for self-oriented, blended learning in medicine: Randomized controlled trial. JMIR Mhealth Uhealth 2017 Sep 14;5(9):e139 [FREE Full text] [doi: 10.2196/mhealth.7943] [Medline: 28912113]
45. Stenlund T, Sundström A, Jonsson B. Effects of repeated testing on short- and long-term memory performance across different test formats. Educ Psychol 2014 Sep 02;36(10):1710-1727. [doi: 10.1080/01443410.2014.953037]

Abbreviations

ANCOVA: analysis of covariance
ICC: intraclass correlation coefficient
IMMS: Instructional Materials Motivation Survey
mAR: mobile augmented reality
m-learning: mobile learning

OS: operating system
POMS: Profile of Mood States
QR: Quick Response
SSL: Secure Sockets Layer
TMD: Total Mood Disturbance
UCEES: Unit of Excellence on Exercise and Health

Edited by G Eysenbach; submitted 20.11.19; peer-reviewed by K Vargas-Román, Y Yilmaz, M Pinheiro; comments to author 12.12.19; revised version received 20.01.20; accepted 26.01.20; published 22.05.20.

Please cite as:
Lozano-Lozano M, Fernández-Lao C, Cantarero-Villanueva I, Noguerol I, Álvarez-Salvago F, Cruz-Fernández M, Arroyo-Morales M, Galiano-Castillo N
A Blended Learning System to Improve Motivation, Mood State, and Satisfaction in Undergraduate Students: Randomized Controlled Trial
J Med Internet Res 2020;22(5):e17101
URL: http://www.jmir.org/2020/5/e17101/
doi:10.2196/17101
PMID:32441655

Original Paper

The Korean Version of the Academic Cyberincivility Assessment Questionnaire for Nursing Students in South Korea: Validity and Reliability Study

Minjoo Hong[1*], PhD; Jennie C De Gagne[2*], PhD, DNP; Hyewon Shin[3*], PhD; Suhye Kwon[4*], PhD; Gum-Hee Choi[5], PhD

[1]Department of Nursing, Gyeongnam National University of Science and Technology, Jinju, Republic of Korea

[2]School of Nursing, Duke University, Durham, NC, United States

[3]School of Nursing, Clemson University, Greenville, SC, United States

[4]College of Nursing, Kosin University, Busan, Republic of Korea

[5]Choonhae College of Health Sciences, Ulsan, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Hyewon Shin, PhD
School of Nursing
Clemson University
605 Grove Road
Greenville, SC, 29605
United States
Phone: 1 864 720 2060
Email: shin@clemson.edu

Abstract

Background: Cybercivility, the practice of what to say and how to say it in online environments, encourages individuals to treat each other with respect. However, the anonymity of online communities may lead some individuals to behave in ways that violate social and cultural norms. These individuals treat others with a lack of regard and even bully others in faceless online confrontations. This practice of cyberincivility can be found across the internet, on commercial sites, and in schools offering online courses. Research on cybercivility and cyberincivility has increased in the United States, where instruments have been developed to measure the impact of cyberincivility in health profession education. However, there is no available instrument that measures nursing students' online behaviors in South Korea.

Objective: The aim of this study was to develop and evaluate a Korean version of the Academic Cyberincivility Assessment Questionnaire developed in the United States.

Methods: Data were collected from 213 nursing students in three South Korean colleges. The Academic Cyberincivility Assessment Questionnaire developed by De Gagne and colleagues was adapted to measure students' knowledge of cybercivility, and their experiences with and acceptability of cyberincivility. Content validity was tested using the content validity index (CVI). Criterion validity was tested using the digital citizenship scale. Reliability was evaluated using Cronbach alpha. The goodness-of-fit of construct validity was determined through exploratory and confirmatory factor analyses.

Results: The CVI was 0.8 or higher for all items. Kuder–Richardson Formula 20, measuring reliability of the knowledge scale, was 0.22 and Cronbach alpha, measuring reliability of the experience scale, was .96. The goodness-of-fit of the model was Chi square=5568.63 (P<.001), the comparative fit index (CFI) was 0.92, and the root mean square error of approximation (RMSEA) was 0.08, which satisfied the criteria. The reliability of the acceptability scale was .96, and the goodness-of-fit indices satisfied the criteria (minimum Chi square/df=2.34, Tucker-Lewis Index =0.92, incremental fit index=0.93, root mean square residual=0.05, CFI=0.93, and RMSEA=0.08).

Conclusions: This study extended and reevaluated the US version of cybercivility scales in a culturally distinct context. The three dimensions of cybercivility include knowledge, experience, and acceptability. Acceptability is well-validated as a dimension, whereas the knowledge dimension requires reexamination for application to Koreans. A revision of the instrument is needed that considers the cultural differences between South Korea and the United States. This paper calls for more attention to be paid to contextualized cybercivility scales among health professions in countries outside the United States.

(J Med Internet Res 2020;22(5):e15668) doi:10.2196/15668

KEYWORDS

cybercivility; health professions education; nursing students; social media; web-based learning

Introduction

Background

Internet and mobile devices are essential to everyday life in the digital age. Most individuals use the internet and social media to keep in touch with friends and family, locate goods and services, and educate themselves on a variety of topics, including health [1]. In South Korea, more than 95% of the population uses smartphones, and the country's use of the internet is among the highest in the world [2]. Moreover, South Korea has an advanced health information technology infrastructure [3]. The use of mobile devices as a means for health information dissemination is well documented [2] and makes South Korea a promising market for such interventions in nursing and nursing education.

The expansion of online networks in which information is exchanged through social networking services (SNSs) has increased opportunities for communication, education, and socialization, but has also led to inappropriate online environment activities [4]. Due to the relative anonymity of individuals meeting online, boundaries are skewed and it is difficult to avoid uncivil exchanges [5,6]. Behaviors that cause harm or unpleasantness in online environments are caused by a lack of awareness as to how opinions, memes, and videos might affect others [7]; a lack of civility in communicating and cooperating with others online; or a lack of literacy that causes well-intentioned posts to convey unintended meanings [8,9]. Cybercivility, defined as the practice of respect and responsibility in the online environment [10], makes it possible to exchange new ideas and make new connections, which is especially important in social networking sites and virtual learning environments.

Cyberincivility is an issue that spans across age and professional groups [1,5]. Recently, research on the behavior of nurses and nursing students revealed a substantial lack of cybercivility, including offensive racial and ethnic remarks through a SNS, posting information without evidence as professional knowledge, using profanity, or violating the confidentiality of patients and coworkers [11,12]. Clinicians or practicum students may breach confidentiality due to lack of civility in online communities by posting patient information via SNSs, providing inaccurate medical information, or posting photos that violate professional ethics [13]. Thus, it is important to provide clear guidelines and education for health profession majors, including nursing students. De Gagne et al [14] developed instruments that measure the knowledge, direct/indirect experience, and perception of cybercivility of health profession students. The authors addressed the issue that future health care professionals experience cyberincivility and perceive incivility as a growing problem. Thus, future health care providers need to be educated on ethical and professional standards, including those that apply in online environments. Although a digital citizenship instrument has been developed to measure incivility in South Korea, since this tool concerns political and global issues [15], Cf it is insufficient to examine the cybercivility of nursing students. Therefore, the aim of the present study was to translate the cyberincivility assessment questionnaire developed in the United States [14] from English into Korean and evaluate its psychometric properties.

Objectives

The aim of this study was to develop a Korean version of the Academic Cyberincivility Assessment Questionnaire (ACAQ) [14] and evaluate its reliability and validity among nursing students in South Korea. These findings could be used as the basis of education about cybercivility for students in health profession education.

Methods

Study Design and Participants

We translated the English version of the ACAQ into Korean and tested the psychometrics of the measures with a survey of nursing students from December 2018 to February 2019. The study participants were students at universities in the Busan and Gyeongnam provinces of South Korea. Approximately 14% of Korean nursing departments are located in these two medium-sized cities in urban locations, which are similar with respect to socioeconomic status. Participants were eligible for inclusion in the study if they were: (1) nursing students and (2) aged 18 years and older. Paper questionnaires consisted of a total of 80 items that included 9 questions about demographics, 15 questions about knowledge, and 28 questions each related to experiences and acceptability of cyberincivility. To perform a factor analysis on the data, the sample size needed to be about five times the number of items in the questionnaire [16]. The appropriate sample size needed for a confirmatory factor analysis (CFA) is at least 200 [17]. The sample size for this study was 213, which satisfied the sample size requirement.

Ethical Considerations

After obtaining approval from the Kosin University Institutional Review Board (KU IRB 2018-0095), we advertised the study on campus. The principal investigator and team members provided paper questionnaires to the researchers or research assistants who explained the research purpose and methods to participants in person. Participation was voluntary, and we explained to the participants that they could drop out at any time during the study. As compensation for their efforts, each participant was given a gift worth the equivalent of 3 USD. We collected 222 surveys, 213 of which were included in the final analysis after excluding 9 surveys due to incomplete information. Permission to use the ACAQ was granted from the original developer of the instrument designed for health profession students (ie, medicine, nursing, physician assistant, and physical therapy) in the United States [14]. Permission from

the authors was also obtained for use of the digital citizenship instrument [15] for assessment of criterion validity of the scale.

Measurements

Questionnaire Development

To assess health profession students' knowledge of cybercivility, and their experience with and perceptions of behaviors related to cyberincivility, De Gagne et al [14] developed the ACAQ. The aim of the present survey was also to determine cybercivility learning needs related to interprofessional education (IPE) core competencies in medical, nursing, physician assistant, and physical therapy programs [14]. The ACAQ consists of 75 items in the following 4 sections: (1) demographics, (2) knowledge about cybercivility, (3) experience and perceptions of cyberincivility, and (4) perceived benefits of including cybercivility education as part of the IPE curriculum. Because the present study surveyed only nursing students and IPE pedagogy is not well-understood among Korean nursing students, we excluded section 4 (perceived benefits of including cybercivility in IPE and preferred formats) from our survey.

Demographics

Demographic questions were based on those in the original studies related to cybercivility [9,14]. The items included in the demographics were gender, year in school, clinical practice experience, SNS memberships (multiple responses), time spent on SNS per day, and number of text messages sent per day.

Knowledge of Cybercivility

The original ACAQ contains 15 items to test students' knowledge of uncivil behaviors in online environments [14]. Response choices were "true," "false," and "I don't know," with a score of 1 and 0 assigned for correct and incorrect answers, respectively; a score of 0 was also assigned for a response of "I don't know." The range of calculated scores was 0-15; the higher the score, the higher the knowledge of cybercivility. The original scale of the reliability measured by Kuder-Richardson formula 20 (KR-20) was 0.58. The scores for KR-20 range from 0 to 1, with a score closer to 1 indicating greater reliability of the test. In general, a score above 0.5 is usually considered to be reasonable [18].

Experience With and Acceptability of Cyberincivility

The original ACAQ includes 28 items in two areas (ie, experience and acceptability) that are responded on a 5-point Likert scale [14]. Respondents were asked to rate how often they had experienced or observed uncivil events described in the questionnaire (1=*never*, 2=*rarely*, 3=*occasionally*, 4=*frequently*, 5=*very often*) and how acceptable they perceived each behavior to be (1=*not at all acceptable*, 2=*slightly acceptable*, 3=*moderately acceptable*, 4=*very acceptable*, 5=*extremely acceptable*). Participants were also asked to report how frequently they experienced or observed their peers, instructors, and other individuals demonstrating certain behaviors. Acceptability was measured by asking the participants to rank the acceptability of the behavior based on its actual or potential consequence(s) related to the students' professional or personal development. Cronbach alpha for the original scale on experience and acceptability with cyberincivility was .95 and .94, respectively.

Digital Citizenship

The digital citizenship instrument developed by Choi and Park [15] to determine the concept of citizenship in the digital age was used as the criterion for evaluating the validity of experience and acceptability of the cyberincivility instrument. This instrument is composed of 23 items in the following five areas evaluated on a 7-point Likert scale: (1) internet political participation, (2) technical ability to use the internet, (3) critical perspective, (4) online communication and collaboration, and (5) sensitivity to community and global issues. The digital citizenship scale has good reliability and construct validity that is supported by an expert panel review, exploratory factor analysis (EFA), and CFA [19]. The digital citizenship scale has been used in several previous studies [19,20]. To determine the reliability of a scale, testing/retesting is necessary, and it should be confirmed that the scales are similar after a retest by applying the same surveys to the same participants after a certain period [21]. However, retesting the same participants would have been difficult and there was the possibility of a decrease in the accuracy due to the prior measurement for this survey. Therefore, we did not use a test-retest method but rather evaluated the validity using the correlations with an existing digital citizenship scale. In a pilot study, there was no ceiling or floor effect observed [22]. After receiving permission from the author, we converted the digital citizenship scale to a 5-point scale to facilitate comparison with the survey of cybercivility. A higher score on this scale indicates a higher level of agreement about digital citizenship. The overall Cronbach alpha was .81 in the original study and was .89 in this study.

Procedure

Instrument Translation and Back Translation

Translation of the instrument was conducted based on World Health Organization guidelines [23] in the order of: (1) preliminary translation, (2) expert panel, (3) back translation, (4) preliminary survey, (5) determination of criterion-related validity, and (6) completion. A bilingual nursing professor working at a nursing college in the United States translated the original instrument into Korean. The translation was confirmed by an investigator who is fluent in both Korean and English. Each item on the scale was verified with the original author for accuracy. For example, item 8 in the *knowledge of cybercivility* section ("Americans encounter incivility almost equally offline and online") was revised with consent of the original author because we assumed cultural differences between South Korea and the United States. Thus, the new statement, "People tend to be ruder online than they are in everyday life," was added to maintain the same number of items in the knowledge section. Similarly, item 14 in the *experience with and perceptions of cyberincivility scale*, "Using displays of attitude such as capitalizing or boldfacing words in an argument," was also modified to "Using new words or abbreviations that seem to mock the other person in discussions in cyberspace" as the Korean language does not have capitalization or bold text. In the expert panel review, there was a comment on possible unfamiliarity of the terms and concepts of cybercivility or

sociocultural differences. Therefore, we reflected on these comments as we translated the instrument. Phrases in the translated instrument were verified by a professor of the Korean language to ensure accuracy. The Korean translation of the experience and acceptability items of the ACAQ is presented in Multimedia Appendix 1.

Preliminary Survey

A preliminary survey was given to 24 nursing students from May 2 to 12, 2018 using the newly translated questionnaire. The purpose of the preliminary survey was to determine the clarity of the content, whether the terminology was easily understood, and if the 15 to 20 minutes allotted to complete the survey was reasonable. The students responded that the survey time was sufficient to answer all questions, and that the items were easy to understand.

Statistical Analysis

The data were analyzed using SPSS 20.0 (IBM Corp, Armonk, NY, USA) and AMOS 22 (IBM SPSS AMOS, Version 22.0. 2013, IBM Corp, Armonk, NY, USA) software. Descriptive statistics, including distribution, were calculated for the demographic variables and item scores. The content validity of the experience and acceptability scales was assessed using the content validity index for individual items (I-CVI) and for scales (S-CVI/Ave). To that end, a 5-member expert panel of four nursing professors and one professor of education scored each item on a 4-point scale (1=not relevant, 2=somewhat relevant, 3=quite relevant, 4=highly relevant).

The I-CVI for each item is computed as the number of experts giving a rating of 3 or 4 divided by the number of experts, and the S-CVI/Ave for the scale is calculated as the mean of the I-CVI values of all items on the scale [21]. The criterion validity of the knowledge section was tested by grouping participants according to whether or not they had taken online classes. To take online classes, students need basic knowledge about and the ability to use the internet [24]; therefore, we assumed that there might be a knowledge gap between those who had taken an online class and those who had not.

The internal consistency and reliability were evaluated using the Cronbach alpha reliability coefficient, which normally ranges between 0 and 1; Cronbach alpha of .8 is considered to be a reasonable goal [25]. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and the Bartlett test of sphericity were conducted to determine the appropriateness of the data for EFA. CFA was used to examine the factor structure. The construct validity was then assessed with EFA using maximum-likelihood estimation with oblique promax rotation. The coefficients for the frequency and acceptability of the cyberincivility items were computed along the eigenvalues of the factors. Each factor was interpreted through examination of the item content, patterns of factor structures, and factor pattern coefficients. Following the EFA, a series of maximum-likelihood estimations via CFA were conducted using AMOS 22.0. The fit of the model was verified by the minimum Chi square/degrees of freedom (CMIN/df) value, comparative fit index (CFI), Tucker-Lewis index (TLI), standardized root mean square residual (SRMR), and root mean square error of approximation (RMSEA). To verify the validity of the criteria, the correlation between experience/acceptability of cyberincivility and digital citizenship was analyzed using Pearson correlation coefficient (r), which ranges from –1 to 1; r of –1 indicates a perfect negative relationship, r of +1 indicates a perfect positive relationship, and r of 0 indicates no linear relationship between variables [25]. The strength of the correlation for the absolute value of r was interpreted as very week for r=0.00-.019, weak for r=0.20-0.39, moderate for r=0.40-0.59, strong for r=0.60-0.79, and very strong for r=0.80-1.0 [26].

Results

Descriptive Statistics

The basic characteristic of the participants are summarized in Table 1. The majority of the participants were women in their 20s who were third-year nursing students with clinical practice experience. The most frequent number of SNS memberships was 1-5, and the most frequently accessed SNS was KakaoTalk, which is a free smartphone app for messaging that is used by most Korean smartphone owners, followed by Facebook, Instagram, Naver Band, KakaoStory, and Twitter. The majority of respondents had taken online classes. Most of the respondents (128/213, 88.3%) with online class experience indicated that the greatest advantage of asynchronistic online classes was the convenience of fitting classes into their personal and professional schedules, while the most common disadvantage reported (101/213, 68.7%) was limited interactions with their professors and peers.

Table 1. Descriptive statistics of survey respondents (N=213).

Characteristic	Value
Age (years), mean (SD)	22.58 (0.78)
Gender, n (%)	
Male	26 (12.2)
Female	187 (87.8)
Year in school, n (%)	
2nd	40 (18.8)
3rd	114 (53.5)
4th	59 (27.7)
Clinical practice experience, n (%)	
Yes	173 (81.2)
No	40 (18.8)
Social networking site memberships, n (%)	
1-5	161 (75.6)
6-10	46 (21.6)
11-20	4 (1.9)
≥21	2 (0.9)
Social network sites (repeated response), n (%)	
KakaoTalk	210 (98.6)
Facebook	187 (87.8)
Instagram	170 (79.8)
Naver Band	117 (54.9)
KakaoStory	109 (51.2)
Twitter	67 (31.5)
Tumblr, Snapchat, WeChat, Flicker, Pinterest, WhatsApp, LinkedIn, and Others	68 (31.8)
Time spent on social network sites daily, n (%)	
Less than 1 hour	20 (9.4)
1-3 hours	122 (57.3)
4-6 hours	61 (28.6)
7-9 hours	9 (4.2)
≥10 hours	1 (0.5)
Experience of online course, n (%)	
Yes	147 (69.0)
No	66 (31.0)

Knowledge of Cybercivility

The average score for cybercivility knowledge was 11.30 (SD 1.86) out of a total of 15 points. The minimum and maximum scores were 5 and 15 points, respectively. Six out of the 15 items were answered correctly by 90% or more of respondents. Three items (items 1, 13, and 14) were correctly answered by 50% or less of the respondents (Table 2). Among them, item 1 received the lowest number of correct answers. Two of these items (items 1 and 14) are related to online privacy protection. The scale used to measure knowledge of cybercivility was a binary scale that assumed possible values of 0 or 1, making it inappropriate for factor analysis. The mean experience measured by frequency was 1.96 (SD 0.78) and the mean perception measured by acceptability was 1.84 (SD 0.72).

Table 2. Participants' knowledge of cybercivility, experience with cyberincivility, and acceptability of cyberincivility (N=213).

Content	Correctly answered, n (%)
An organization ensures that all information it collects about users will be kept confidential.	54 (25.4)
Cyberbullying is a form of incivility that occurs in cyberspace where online communication happens.	203 (95.3)
Cyberincivility is a concern among general college populations, but it has nothing to do with students' learning outcomes.	153 (71.8)
Cyberincivility occurs in social media channels, online learning environments, and email.	195 (91.5)
Ethical standards guiding appropriate use of social media and online networking forums in education are already well-established.	134 (62.9)
People say and do things online that they would not say or do in person.	202 (94.8)
Posting unprofessional content online can reflect unfavorably on health profession students, faculty, and institutions.	192 (90.1)
People tend to be ruder online than they are in everyday life. (Original: Americans encounter incivility almost equally offline and online)	206 (96.7)
Unlike traditional bullying, cyberbullying does not require repeated behavior.	184 (86.4)
Cyberincivility is linked to higher stress levels, lower morale, and incidences of physical harm.	199 (93.4)
Using social media inappropriately cannot lead to civil or criminal penalties.	160 (75.1)
Cyberincivility does not occur in the workplace.	184 (86.4)
Humor, anger, and other emotional components of online messages are the same as face-to-face messages.	98 (46.0)
Breaches of confidentiality on social media may lead to mandatory reporting to licensing and credentialing bodies.	113 (53.1)
Despite privacy settings on social media, nothing is private after it is posted on the internet.	129 (60.6)

The I-CVI of cybercivility knowledge was more than 0.80 and the S-CVI/Ave was 0.92. The criterion validity of the knowledge scale was tested using the known-groups technique. The knowledge of cybercivility among those who took at least one online class was 11.38 (SD 1.85) and was 11.26 (SD 1.87) for those who did not take online classes, which did not differ significantly (t_{125}=0.437, P=.96). The reliability of the knowledge scale assessed by KR-20 was 0.22.

Experience With Cyberincivility

Content Validity

The average score for experience with cyberincivility was 1.96 (SD 0.78) out of 5 points. The I-CVI of cybercivility experience was more than 0.80 and S-CVI/Ave was 0.98.

Exploratory Factor Analysis

The KMO measure of sampling adequacy for experience with cyberincivility yielded an index of 0.94. Bartlett test of sphericity was significant (Chi square=5568.63, P<.001), indicating that the data were appropriate for EFA. When the EFA was run using maximum-likelihood estimation with oblique promax rotation, 4 factors with eigenvalues≥1.0 were extracted and accounted for 72.22% of the overall variance. The factor analysis showed that the commonalities of all items were 0.40 or higher and the eigenvalues were 1 or greater, resulting in 4 factors being extracted. The first item was "Blaming technology for failure of communication, assignment completion, or submissions." This was excluded from the EFA because it was not included in any factor. CFA was then performed using the 4 factors extracted in the EFA. We named these four factors F1 (individual behaviors in online environments), F2 (online class attendance attitude), F3 (email manner in online environments), and F4 (online assignment ethics).

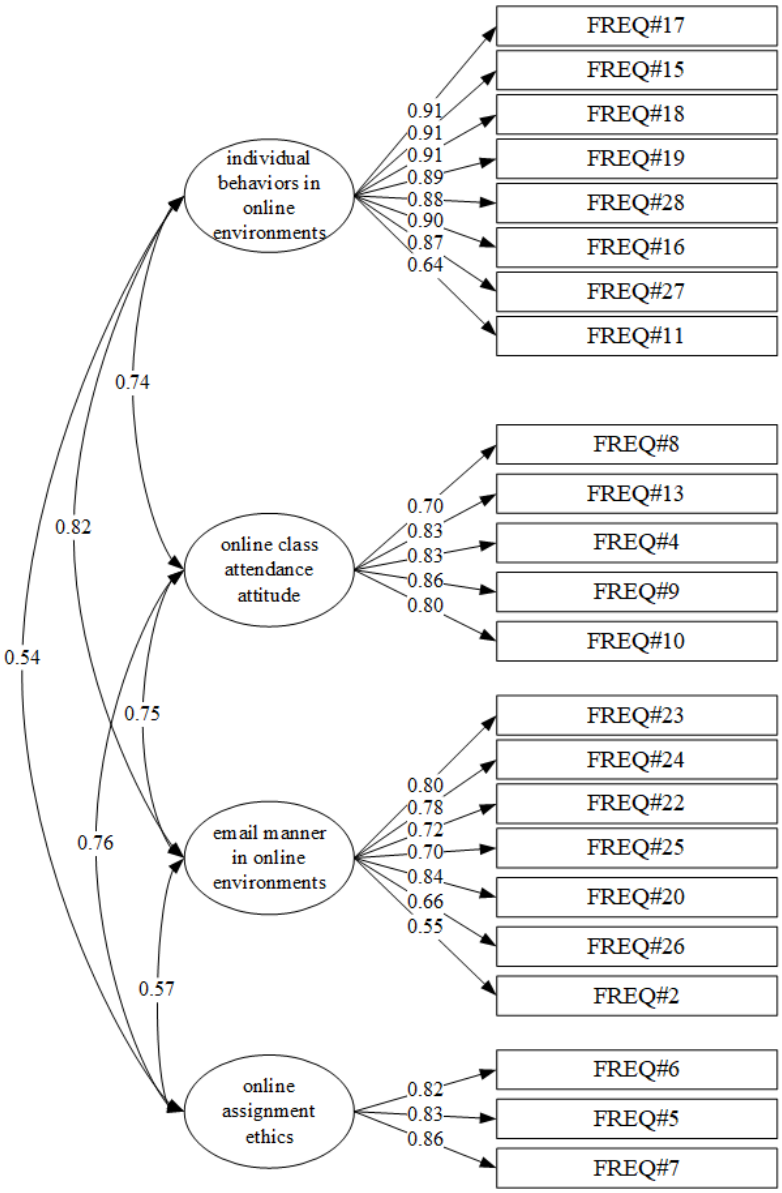
Confirmatory Factor Analysis

A CFA was performed on the 4 subfactor models extracted via the EFA. The path models are shown in Figure 1. The goodness-of-fit indices of the 4-factor structural model were as follows: CMIN/DF=2.724, TLI=0.90, incremental fit index (IFI)=0.92, SRMR=0.06, CFI=0.92, and RMSEA=0.08. The best-fit model criteria were CMIN/df≤3, TLI≥0.90, IFI≥0.90, SRMR≤0.08, CFI≥0.90, and RMSEA≤0.08, which suggests that all of the goodness-of-fit indices of the model satisfied the criteria.

Reliability

Cronbach alpha coefficient of experience with cyberincivility was .96, and the Cronbach alpha coefficients of the 4 subareas were .96 for F1 (individual behaviors in online environments), .90 for F2 (online class attendance attitude), .88 for F3 (email manner in online environments), and .88 for F4 (online assignment ethics).

Figure 1. Confirmatory factor analysis of experience of cyberincivility.



Criterion Validity

The content validity of experience was good, but the criterion validity was low. The experience of cyberincivility [14] was calculated using Pearson correlation coefficient, which was not significant (r=0.085, P=.22).

Acceptability of Cyberincivility

Content Validity

The average score for acceptability of cyberincivility was 1.84 (SD 0.72) out of 5 points. Criterion validity was tested by comparing the means of correlations of the digital citizenship instrument [15].

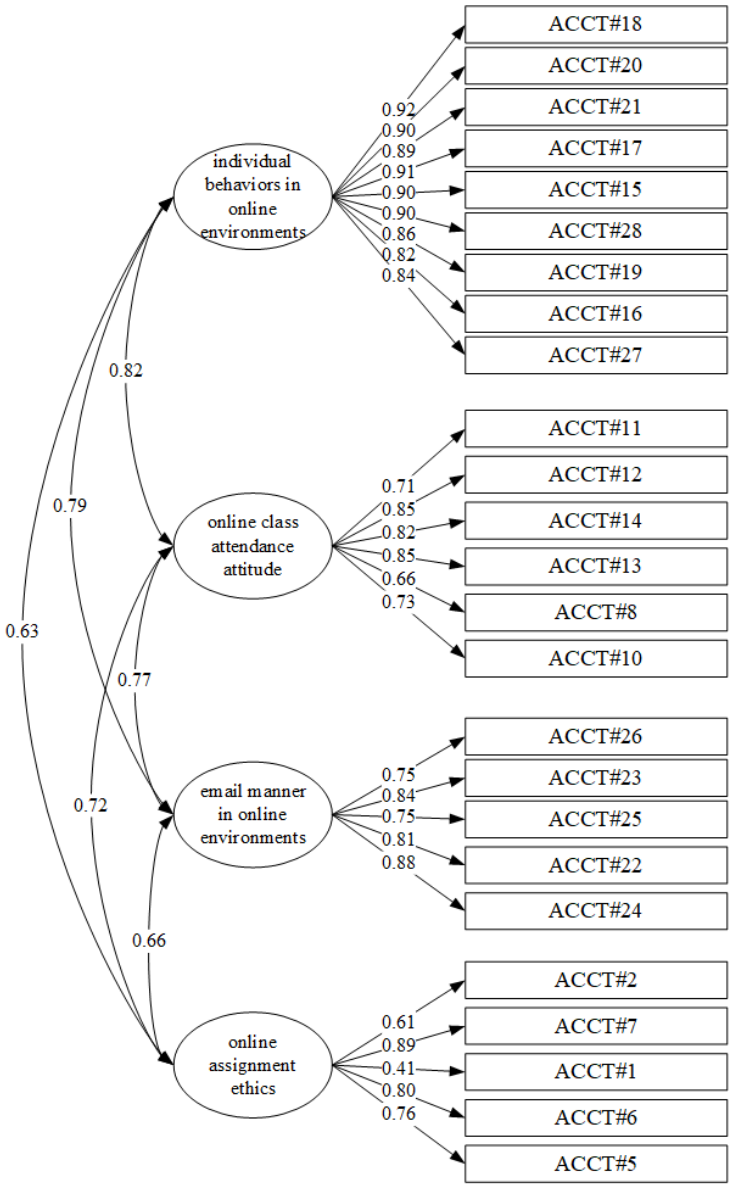
Exploratory Factor Analysis

The KMO measure of sampling adequacy for acceptability of the cyberincivility items yielded an index of 0.94. Bartlett test of sphericity was significant (Chi square=5635.51, P<.001), indicating that the data were appropriate for EFA. When the EFA was run using maximum-likelihood estimation with oblique promax rotation, the commonalities of all items were 0.40 or higher, and the eigenvalues were 1 or greater, resulting in 4 factors being extracted. The overall explanatory power of the analysis was 70.53%, and all values conformed to the goodness-of-fit criteria.

Confirmatory Factor Analysis

A CFA was performed on the 4-factor model based on the 4 subfactors extracted in the EFA; the path model is presented in Figure 2. The goodness-of-fit indices of the 4-factor structural model were as follows: CMIN/df=2.343, TLI=0.92, IFI=0.93, SRMR=0.05, CFI=0.93, and RMSEA=0.08. All of the goodness-of-fit indices of the model satisfied the criteria.

Figure 2. Confirmatory factor analysis of acceptability of cyberincivility.



Reliability

Cronbach alpha coefficient of the acceptability of cyberincivility was .96. Cronbach alpha coefficients of the subcategories were .97 for F1 (individual behaviors in online environments), .89 for F2 (online class attendance attitude), .90 for F3 (email manner in online environments), and .81 for F4 (online assignment ethics).

Criterion Validity

Criterion validity was tested by means of correlations with the digital citizenship instrument [15]. The acceptability of cyberincivility was tested using Pearson correlation coefficient, which was statistically significant (r=0.16, P<.001).

Discussion

Principal Findings

A Korean version of the cyberincivility assessment questionnaire was developed to test knowledge of cybercivility and to explore the experiences and acceptability of uncivil behavior in various cyber domains. The translated questionnaire was then used to measure the knowledge, experiences, and acceptability of nursing students.

Most of the participants of this study were female students who had completed 2 years toward their degrees, a nursing practicum, and at least one online lecture. Since female students make up the majority of nursing students, 80% of the participants were female. In the original study in the United States in which 205 students participated, most of them were also female at 83.4% [14]. Comparing the two studies, 75% of the participants in the current study and 82% of the US participants [14] reported that they were members of 1-5 SNSs.

The content of the Korean scale was deemed to be appropriate to measure the knowledge of cybercivility, experience, and acceptability of cyberincivility as shown in the CVI. Polit et al [21] recommended that for a scale to be judged as having excellent content validity, it should comprise items with I-CVIs of 0.78 or higher for three or more experts and an S-CVI/Ave

of 0.90 or higher. The content validity using I-CVI and S-CVI/Ave of the knowledge, experiences, and acceptability scales satisfied these standards. This might reflect sufficient reviews during the translation by scholars and professors of a Korean expert in second languages.

The average score for cybercivility knowledge was similar to the result of the original study (11.30, SD 1.86 vs 11.53, SD 1.99, respectively) [14]. However, the current study's reliability of cybercivility knowledge was lower than that observed in the original study. The low rate of correct answers may have been due to participants' lack of knowledge or exposure to certain types of cyberactivity. The fact that traditional lectures are more common in South Korea might have been a factor contributing to this difference. In addition, faculty tend to use one-way communication in asynchronistic online classes, such as uploading lectures with voice recordings instead of holding discussions on online portals or conducting virtual online meetings with students in South Korea. Indeed, in a study that examined 2600 courses at one university in South Korea, only 3% of online courses used internet-based discussions and only 8% used online meeting programs to facilitate group work [27]. In addition, the participants of this study were undergraduate students with a mean age of 22.58 (SD 0.78) years, whereas the subjects of the original study were graduate students with a mean age of 29.70 (SD 7.0) years [14]. This age difference might have led to a difference in their experience and knowledge and caused the variation in their responses. It is recommended that future studies consider these results and reconfirm KR-20 with a greater variety of subjects.

The digital citizenship instrument [15] was used to test the criterion-related validity with the translated cyberincivility assessment instrument. The differences between the tests were likely related to their different contexts and populations. That is, the digital citizenship instrument was composed of items designed to address general social relationships on internet use or internet political participation, whereas the cyberincivility assessment questionnaire was developed to address the academic environment from the perspective of students in health-related majors [14]. However, due to differences in communication methods or meanings in different sociocultural environments [28], the definition of cybercivility can vary among individuals and countries. In the case of the United States, it is common to use email for personal and business exchanges and to share opinions in learning management systems during online classes or traditional classes with this online component [29], so that students in a US academic setting are more accustomed to these methods of communication. However, in South Korea, text messaging services using mobile phones are the most frequently used channel of communication among students and faculty [30]. Hence, it is not surprising that differences were found in what constitutes civil or uncivil behavior between these two populations. Furthermore, communication methods and manners in text messaging services are more of a concern than emails or online college communication tools [27].

The significance of the present study lies in the validity of an instrument that can measure the cybercivility knowledge and practices of South Korean nursing students whose professionalism and attitudes are important to their careers and patient care. For future study, to better understand an individual's experience of cyberincivility, there is a need for measuring an individual's direct experiences in addition to observations of others' behaviors in online environments. It is also necessary to modify the knowledge cybercivility assessment questionnaire by considering cultural aspects to obtain a good reliability coefficient. The ways in which individuals communicate and their manners in cyberspace differ between South Korea and the United States, necessitating a revision of the instrument to take into account cultural differences and to provide greater validity for future studies.

Limitations

The experience scale is limited because findings are based on self-reports of students' personal experiences and their observations of others. Although the correlation between experience/acceptability of cyberincivility and the digital citizenship score was significant, care should be exercised in interpreting the results because the instrument might not measure exactly what it was intended to measure due to the low correlation. In addition, owing to cultural differences related to cybercivility, an English instrument rewritten in Korean might have resulted in some words getting lost in translation. The same difficulty would be true for translations into any other languages. Finally, when participants respond to a survey, they tend to give socially favorable responses, which is even more evident for Koreans than Westerners [31].

Conclusions

This study extended and reevaluated the US version of cybercivility scales in a culturally distinct context. The results show that among the three dimensions of cybercivility (knowledge, experience, and acceptability), acceptability is well-validated while the knowledge dimension requires reexamination in the South Korean context. The content validity of experience was good, but the criterion validity was low. This study calls for more attention to contextualized cybercivility scales among health professions in countries outside the United States.

Acknowledgments

This work was supported by a National Research Foundation of Korea grant from the Ministry of Science, ICT and Future Planning of the Korean Government (No. 2018R1C1B5086516).

Authors' Contributions

HM, JCD, SH, KS, and CG were responsible for conceptual development, research design, and drafting of the manuscript; HM, SG, and GC were responsible for collecting data and data interpretation. All authors discussed the results and contributed to the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Korean translation of the experience and acceptability items of the Academic Cyberincivility Assessment Questionnaire.
[DOCX File , 20 KB - jmir_v22i5e15668_app1.docx]

References

1. Norman CD. Social media and health promotion. Glob Health Promot 2012 Dec;19(4):3-6. [doi: 10.1177/1757975912464593] [Medline: 24803437]
2. Poushter J, Bishop C, Chwe H. Pew Research Center. 2018. Social media use continues to rise in developing countries but plateaus across developed ones URL: https://tinyurl.com/ya936vhh [accessed 2019-05-21]
3. Chung S, Park BK, Nahm E. The Korean eHealth Literacy Scale (K-eHEALS): Reliability and Validity Testing in Younger Adults Recruited Online. J Med Internet Res 2018 Apr 20;20(4):e138 [FREE Full text] [doi: 10.2196/jmir.8759] [Medline: 29678800]
4. Kim T, Kang M. Type Analysis and Countermeasures of Side Effects of using Smart Phone. J Korean Inst Inform Commun Eng 2013 Dec 31;17(12):2984-2994. [doi: 10.6109/jkiice.2013.17.12.2984]
5. Suler JR. Psychology of the Digital Age. New York: Cambridge University Press; 2019.
6. Yoon KY. Citizenship and moral education in the age of cyber communication. Korea Elementary School Education for Moral Education 2015;581:581-584.
7. Lee JK, Woo HJ. A study on the intention of cyber language violence. Cyber Commun J 2010;27(1):215-253.
8. Choi MS, Hyung JP. Variables affecting the digital citizenship of college students. Citizen Educ Res 2016;48(3):211-237. [doi: 10.35557/trce.48.3.201609.008]
9. Winn MR. Learning & Leading with Technology.: International Society for Technology Information; 2012. Promote digital citizenship through school-based social networking URL: https://files.eric.ed.gov/fulltext/EJ954323.pdf [accessed 2019-05-21]
10. Kim YI. Experience of youth participation in youth council in Korea and changes in citizenship. Youth Studies 2007;14(6):99-127.
11. De Gagne JC, Choi M, Ledbetter L, Kang HS, Clark CM. An Integrative Review of Cybercivility in Health Professions Education. Nurse Educ 2016;41(5):239-245. [doi: 10.1097/NNE.0000000000000264] [Medline: 27022683]
12. De Gagne JC, Yamane SS, Conklin JL. Evidence-based strategies to create a culture of cybercivility in health professions education. Nurse Educ Today 2016 Oct;45:138-141. [doi: 10.1016/j.nedt.2016.07.007] [Medline: 27472429]
13. De Gagne JC, Hong M. Do you have cybercivility?. Seoul; 2019. The Korean Nurse Association News URL: http://www.nursenews.co.kr/main/ArticleDetailView.asp?sSection=100&idx=23992&intPage=1 [accessed 2019-05-21]
14. De Gagne JC, Conklin JL, Yamane SS, Kang HS, Hyun M. Educating Future Health Care Professionals About Cybercivility: Needs Assessment Study. Nurse Educ 2018;43(5):256-261. [doi: 10.1097/NNE.0000000000000487] [Medline: 29206713]
15. Choi MS, Park H. A Validation Study on the Korean Digital Citizenship Scale through Exploratory and Confirmatory Factor Analysis. Theory Res Citizenship Educ 2015 Dec;47(4):273-297. [doi: 10.35557/trce.47.4.201512.009]
16. Tabachnick B, Fidell L, Ullman J. Using multivariate statistics, 6th edition. Boston: Pearson; 2013.
17. Tak J. Psychological testing: An understanding of development and evaluation method. 2nd Ed. Seoul: Hakjisa; 2007.
18. McGahee TW, Ball J. How to Read and Really Use an Item Analysis. Nurse Educator 2009;34(4):166-171. [doi: 10.1097/nne.0b013e3181aaba94] [Medline: 19574855]
19. Choi M, Glassman M, Cristol D. What it means to be a citizen in the internet age: Development of a reliable and valid digital citizenship scale. Comput Educ 2017 Apr;107:100-112. [doi: 10.1016/j.compedu.2017.01.002]
20. Choi M. A Concept Analysis of Digital Citizenship for Democratic Citizenship Education in the Internet Age. Theory Res Social Educ 2016 Aug 25;44(4):565-607. [doi: 10.1080/00933104.2016.1210549]
21. Polit DF, Beck CT, Hungler BP. Essentials of nursing research: methods, appraisal, and utilization 5th ed. Philadelphia, PA: Lipincott, Williams and Wilkins; 2001.
22. Stucki G, Liang MH, Stucki S, Katz JN, Lew RA. Application of statistical graphics to facilitate selection of health status measures for clinical practice and evaluative research. Clin Rheumatol 1999 Apr 1;18(2):101-105. [doi: 10.1007/s100670050065] [Medline: 10357113]
23. Word Health Organization. Process of translation and adaptation of instruments URL: http://www.who.int/substance_abuse/research_tools/translation/en [accessed 2018-03-13]

24. Kim BC. A Q-study on the subjective perception types in users of Internet newspaper ads. J Cybercommun Acad Soc 2005;16:345-374.
25. Osborn CE. Statistical Applications for Health Information Management. Burlington, MA: Jones & Barlett Learning; 2005.
26. Evans JD. Straightforward Statistics for the Behavioral Sciences. Pacific Grove, CA: Duxbury Press; 1995.
27. Park Y, Yu JH, Jo I. Clustering blended learning courses by online behavior data: A case study in a Korean higher education institute. Internet Higher Educ 2016 Apr;29:1-11. [doi: 10.1016/j.iheduc.2015.11.001]
28. Bae JH. A comparative analysis between Korea and America in terms of the effects of mobile phone use on interpersonal networks: Focusing on college student populations. J Commun Sci 2009;9(3):178-210.
29. Lindsay M. Inside Higher Education. 2018 Jul 10. Canvas catches, and maybe passes, Blackboard URL: https://www.insidehighered.com/digital-learning/article/2018/07/10/canvas-catches-and-maybe-passes-blackboard-top-learning [accessed 2019-04-19]
30. Koh YW, Sohn EM, Lee HJ. Users' perception on fonts as a tool of communication and SMS. J Korean Soc Design Sci 2007;20(1):133-142.
31. Kim SH, Kim S. National Culture and Social Desirability Bias in Measuring Public Service Motivation. Admin Soc 2013 Sep 04;48(4):444-476. [doi: 10.1177/0095399713498749]

Abbreviations

ACAQ: Academic Cyberincivility Assessment Questionnaire
CFA: confirmatory factor analysis
CFI: comparative fit index
CMIN: minimum Chi square
CMIN/df: minimum Chi square/degrees of freedom
CVI: content validity index
EFA: exploratory factor analysis
I-CVI: content validity index for individual items
IFI: incremental fit index
IPE: interprofessional education
KMO: Kaiser-Meyer-Olkin
KR-20: Kuder-Richardson formula 20
RMSEA: root mean square error of approximation
S-CVI/Ave: content validity index for scales
SNS: social networking service
SRMR: standardized root mean square residual
TLI: Tucker-Lewis index

Edited by G Eysenbach; submitted 27.07.19; peer-reviewed by BK Park, H Hah, J Parsons, R Haase, MA Mayer; comments to author 19.08.19; revised version received 23.10.19; accepted 10.04.20; published 05.05.20.

Please cite as:
Hong M, De Gagne JC, Shin H, Kwon S, Choi GH
The Korean Version of the Academic Cyberincivility Assessment Questionnaire for Nursing Students in South Korea: Validity and Reliability Study
J Med Internet Res 2020;22(5):e15668
URL: https://www.jmir.org/2020/5/e15668
doi:10.2196/15668
PMID:32369027

Original Paper

Transfer Learning for Risk Classification of Social Media Posts: Model Evaluation Study

Derek Howard[1,2], MSc; Marta M Maslej[1,2], PhD; Justin Lee[3], BSc; Jacob Ritchie[1,4], MSc; Geoffrey Woollard[5,6], MSc; Leon French[1,2,7,8], PhD

[1]Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, ON, Canada
[2]Krembil Centre for Neuroinformatics, Centre for Addiction and Mental Health, Toronto, ON, Canada
[3]Department of Biochemistry, University of Toronto, Toronto, ON, Canada
[4]Department of Computer Science, University of Toronto, Toronto, ON, Canada
[5]Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada
[6]Princess Margaret Cancer Centre, University Health Network, Toronto, ON, Canada
[7]Institute for Medical Science, University of Toronto, Toronto, ON, Canada
[8]Division of Brain and Therapeutics, Department of Psychiatry, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Leon French, PhD
Krembil Centre for Neuroinformatics
Centre for Addiction and Mental Health
250 College St
Toronto, ON, M5T 1R8
Canada
Phone: 1 (416) 535 8501
Email: leon.french@camh.ca

Abstract

Background: Mental illness affects a significant portion of the worldwide population. Online mental health forums can provide a supportive environment for those afflicted and also generate a large amount of data that can be mined to predict mental health states using machine learning methods.

Objective: This study aimed to benchmark multiple methods of text feature representation for social media posts and compare their downstream use with automated machine learning (AutoML) tools. We tested on datasets that contain posts labeled for perceived suicide risk or moderator attention in the context of self-harm. Specifically, we assessed the ability of the methods to prioritize posts that a moderator would identify for immediate response.

Methods: We used 1588 labeled posts from the Computational Linguistics and Clinical Psychology (CLPsych) 2017 shared task collected from the Reachout.com forum. Posts were represented using lexicon-based tools, including Valence Aware Dictionary and sEntiment Reasoner, Empath, and Linguistic Inquiry and Word Count, and also using pretrained artificial neural network models, including DeepMoji, Universal Sentence Encoder, and Generative Pretrained Transformer-1 (GPT-1). We used Tree-based Optimization Tool and Auto-Sklearn as AutoML tools to generate classifiers to triage the posts.

Results: The top-performing system used features derived from the GPT-1 model, which was fine-tuned on over 150,000 unlabeled posts from Reachout.com. Our top system had a macroaveraged F1 score of 0.572, providing a new state-of-the-art result on the CLPsych 2017 task. This was achieved without additional information from metadata or preceding posts. Error analyses revealed that this top system often misses expressions of hopelessness. In addition, we have presented visualizations that aid in the understanding of the learned classifiers.

Conclusions: In this study, we found that transfer learning is an effective strategy for predicting risk with relatively little labeled data and noted that fine-tuning of pretrained language models provides further gains when large amounts of unlabeled text are available.

(J Med Internet Res 2020;22(5):e15371) doi:10.2196/15371

KEYWORDS

triage; classification; natural language processing; transfer learning; machine learning; data interpretation, statistical; mental health; social support

Introduction

Mental health disorders are highly prevalent, with epidemiological studies reporting roughly half the population in the United States meeting the criteria for one or more mental disorders in their lifetime and roughly a quarter meeting the criteria in a given year [1]. Available survey evidence suggests that the first onset of mental health disorders is typically in childhood or adolescence and that later-onset disorders are mostly secondary conditions. The severity of mental disorders is highly related to their comorbidity, with complex interactions among disorders [2]. Moreover, severe disorders tend to be preceded by less severe disorders that are often not brought to clinical attention, indicating a need for early detection and intervention strategies [3,4].

Mental disorders are among the strongest predictors for nonsuicidal self-injury and suicidal behaviors; however, little is known about how people transition from suicidal thoughts to attempts [5]. Given the high incidence of mental health disorders and the relatively low incidence of suicide attempts, predicting the risk for suicidal behavior is difficult. In particular, Franklin et al [6] report a lack of progress over the last 50 years on the identification of risk factors that can aid in the prediction of suicidal thoughts and behaviors. However, they also proposed that new methods with a focus on risk algorithms using machine learning present an ideal path forward. These approaches can be integrated into peer support forums to develop repeated and continuous measurements of a user's well-being to inform early interventions.

Peer support forums can be a useful and scalable approach to social therapy for mental health issues [7]. Many individuals are already seeking health information online, and this manner of information access can help those who are reluctant to seek professional help, are concerned about stigma or confidentiality, or face barriers to access [8]. There is limited evidence showing that online peer support without professional moderation is an effective strategy for enhancing users' well-being [7,9]. However, in a systematic review of social networking sites for mental health interventions, Ridout and Campbell [10] identified the use of moderators as a key component of successful interventions on these Web-based platforms. The development of automated triage systems in these contexts can facilitate professional intervention by prioritizing users for specialized care [11,12] or decreasing response time when a risk for self-harm is identified [13]. Although the computational infrastructure of peer support forums is scalable, the effectiveness of human moderation is challenging to grow with community size. If they are accurate, automated systems can address these needs through computational approaches that are fast and scalable.

Previous research suggests that the language of individuals with mental health conditions is characterized by distinct features [14-17], eg, frequent use of first-person singular pronouns has been associated with depression [18]. This has sparked efforts to develop automated systems that, when given social media data, can predict the same level of suicide or self-harm risk that a trained expert would predict.

Such automated systems typically start with a feature extraction step that converts the variable length input text into fixed-length numeric vectors (features). This step is required to apply machine learning classifiers that operate on such vectors. An example is the bag-of-words representation, where each numeric feature represents the count of a specific word that is selected based on frequency or from a lexicon. With such a representation, a classifier may learn that mentions of *hopeless* are more common in text written by depressed individuals. This step of extracting features that best represent the text is a key part of such systems because a significant amount of information loss can occur. For example, in the bag-of-words representation, the order of the words is discarded. In contrast, differences in performance across machine learning classifiers are lower when representations are held constant. For example, good classifiers will have a similar performance on the same representations. Lexicon-based tools are highly dependent on their dictionaries, which require manual curation and validation. However, lexicon- and rule-based approaches are typically more interpretable than more complex neural network–based representations.

Recently, word embeddings have been shown to provide rich representations where words from the same context of a corpus tend to occupy a similar feature space [19]. The use of these embeddings has significantly boosted performance in several natural language processing tasks in recent years [20]. Generating such word embeddings can be done by building a neural network model that predicts a word given its neighboring words or vice versa. These word representations are learned from large corpora. These representations can be reused for other tasks. For example, a pretrained representation of *hopeless* would be similar to *despair*, allowing a classifier to group text that shares these words. This reuse is a type of transfer learning, which allows for the knowledge learned from one domain to be transferred to a task in an adjacent domain [21]. More recently, pretrained word representations have been shown to capture complex contextual word characteristics better than the preceding shallow models [22]. The fine-tuning of large pretrained language models in an unsupervised fashion has pushed forward the applicability of these approaches in cases with small amounts of labeled data [20,23]. Such fine-tuning could alter the learned context of *worries* to account for its placement in the common Australian expression of *no worries* when being transferred from an American to Australian corpus. Given these recent advances in natural language processing, we tested the performance of transfer learning with pretrained language models on risk classification of social media posts.

Reachout.com is an Australian youth-based mental health peer support forum. It is targeted for those aged 14 to 25 years, and the community is maintained by staff and trained volunteer

moderators. Staff and moderators monitor the forums, and they respond, as required, with empathy, support, and referrals to relevant information and available services. The 2017 Computational Linguistics and Clinical Psychology (CLPsych)–shared task organizers provided a corpus of posts from Reachout.com to assess the ability of automated methods to triage forum posts based on the urgency of moderator response [24]. For example, posts that suggest the author might hurt themselves or others are labeled as being high in priority for moderator response (*crisis*). We noted that these labels do not distinguish if the author is contemplating self-harm, nonsuicidal self-injury, or suicide. These constructs have different prevalence and incidence rates [25]. This dataset is small and imbalanced as the majority of posts are labeled as not requiring a moderator response. For example, only 5.2% (82/1588) of the posts are labeled as *crisis*. Given the higher importance of posts requiring moderator response, the organizers of the CLPsych-shared task chose the macroaveraged F1 metric to weight performance equally across the labels that mark the urgency of moderator response. This metric weighs each of those labels for both precision and recall equally. As a result, misclassification of a *crisis* post will be costlier because *crisis* posts occur less frequently. Several advanced methods have been applied to this dataset [24,26], but a systematic evaluation of feature extraction methods has not been performed.

In this paper, we benchmarked multiple feature extraction methods on forum posts from Reachout.com by evaluating their ability to predict the urgency of moderator response. Furthermore, we explored the interpretability through emoji representations and by visualizing word importance in text that mimics themes from suicide notes. We have shown that modern transfer learning approaches that take advantage of large corpora of unlabeled text, in combination with automated machine learning (AutoML) tools, improve performance.

Methods

Data

Reachout.com

Our primary data source was made available for the 2017 CLPsych-shared task and was collected from the Australian mental health peer support forum, Reachout.com [13,24]. The entire dataset consisted of 157,963 posts written between July 2012 and March 2017. Of those, 1188 were labeled and used for training the classification system, and 400 labeled posts were held out for the final evaluation of the systems. Posts were labeled *green* (58.6%, 931/1588), *amber* (25.6%, 390/1588), *red* (11.7%, 185/1588), or *crisis* (5.2%, 82/1588) based on the level of urgency with which moderators should respond. The postannotation task began with the 3 judges (organizers of the shared task) discussing and coming to a shared agreement on the labels for roughly 200 posts, guided by informal annotation and triage criteria provided by Reachout. The annotators ultimately formalized their process in a flowchart to standardize the labeling process and included fine-grained or granular annotations for each of the posts (Summary table of fine-grained labels in Multimedia Appendix 1). They then annotated the remaining posts independently, and the interannotator agreement was measured over these posts (excluding 22 posts labeled ambiguous by at least one judge). The 3 judges achieved a Fleiss kappa of 0.706 and a pairwise Cohen kappa score ranging from 0.674 to 0.761, which is interpreted as substantial agreement by Viera and Garrett [27]. The above mentioned steps, evaluations, and development of this dataset were previously undertaken by Milne et al [13,24].

University of Maryland Reddit Suicidality Dataset

To test the generalizability of the system developed on the Reachout.com data, we used a subset of the data made available from the University of Maryland (UMD) Reddit Suicidality Dataset [28,29]. The collection of this dataset followed an approach where the initial signal for a positive status of suicidality was a user having posted in the subreddit, /r/SuicideWatch, between 2006 and 2015. Annotations were then applied at the user level based on their history of posts. We used the subset that was curated by expert annotators to assess suicide risk. These volunteer experts include a suicide prevention coordinator for the Veteran's Administration; a cochair of the National Suicide Prevention Lifelines Standards, Training, and Practices Subcommittee; a doctoral student with expert training in suicide assessment and treatment whose research is focused on suicidality among minority youths; and a clinician in the Department of Emergency Psychiatry at Boston Children's Hospital. Two sets of annotator instructions (short and long) were used, following an adapted categorization of suicide risk developed by Corbitt-Hall et al [30]: (a) no risk (or *None*): I don't see evidence that this person is at risk for suicide, (b) low risk: There may be some factors here that could suggest risk, but I don't really think this person is at much of a risk of suicide, (c) moderate risk: I see indications that there could be a genuine risk of this person making a suicide attempt, and (d) severe risk: I believe this person is at high risk of attempting suicide in the near future. These categories correspond roughly to the green, amber, red, and crisis categories defined in the Reachout.com data. The longer set of annotation instructions also identified 4 families of risk factors (ie, thoughts, feelings, logistics, and context). A pairwise Krippendorff alpha was used to assess interannotator agreement, with an average alpha of .812 satisfying the recommendation of a reliability cutoff of alpha >.800 [31]. Consensus labels were determined using a model for inferring true labels from multiple noisy annotations [32,33]. The abovementioned steps and development of this dataset were undertaken by Shing and et al [29].

Of the subset with labels by expert annotators, we then selected only data from users who had posted once in /r/SuicideWatch to minimize ambiguity in understanding which of their posts was the cause of the associated label. Predictions were made only on posts from /r/SuicideWatch. In total, there were 179 user posts across the categories (*a*: 32, *b*: 36, *c*: 85, and *d*: 26). The Centre for Addiction and Mental Health Research Ethics Board approved the use of this dataset for this study.

To better gauge our performance on the UMD Reddit Suicidality Dataset posts, we calculated an empirical distribution of random baselines for the macro-F1 metric. This baseline distribution quantifies the performance of random shuffles of the true labels (including the class *a* or *no risk* labels). As expected, across

10,000 of these randomizations, the mean macro-F1 was 0.25. We set a threshold of 0.336, which is 62 of 10,000 random runs to mark Reddit validation performance as better than chance (1/20 × 1/8 × 10000), corresponding to $P<.05$ and a Bonferroni correction for 8 tests (number of feature sets tested).

Composite Quotes

We used 10 composite quotes to share example predictions of our system on text that could be predictive/indicative of self-harming and/or suicidality. These composite quotes were created by Furqan et al [34] and were derived from qualitative research that synthesized primary themes noted in a selection of suicide notes that made explicit mentions of mental illness or mental health care. To assess the role of individual words (or tokens) in the classification of the quote, we iteratively perturbed each token and replaced it with an unknown token outside of the model's vocabulary and reran the prediction.

Data Preprocessing and Feature Extraction

Features were extracted from only the text body of the posts. For all posts, any quotes from previous posts or links to images were removed.

We extracted features using lexicon-based tools such as Valence Aware Dictionary and sEntiment Reasoner (VADER; 4 features) [35], Linguistic Inquiry and Word Count (LIWC; 70 features) [36], and Empath (195 features) [37], which have proven to be useful for characterizing social media text and extracting psychologically relevant signals. Features were also extracted from 3 pretrained artificial neural network models: DeepMoji [38] was used to extract sentiment- and emotion-related features (eg, the use of emoticons in social media text), the Universal Sentence Encoder version 2 (using a deep averaging network encoder) (Google) [39] obtained from Tensorflow Hub that was specifically designed to facilitate transfer learning, and the Generative Pretrained Transformer (GPT) network version 1 (OpenAI) [20]. For DeepMoji, we extracted features that represent the 64 predicted emojis and the neural activations from the preceding attention layer in the network (2304 features, referred to as DeepMoji). We used the Indico Data Solutions implementation to extract features from the default pretrained GPT-1 network and also after fine-tuning on the unlabeled corpus of posts from Reachout.com [40]. All language model fine-tuning was done with 3 epochs over the unlabeled posts, as suggested by the GPT-1 authors.

With Empath and LIWC, sentence splitting was not performed. With the remaining feature encoding (VADER, DeepMoji, Universal Sentence Encoder, and both GPT models) methods, we first preprocessed the text body of each post into sentences using the sentence boundary detection from spaCy version 2.1. Sentence feature vectors were aggregated to the post level by taking their mean, maximum, and minimum for each extracted feature.

Model Optimization and Selection

To train classifiers on the various feature sets, we used 2 AutoML methods that are built upon scikit-learn [41] to optimize and select optimal models. We selected these tools over others because they are open source. Other AutoML tools may have advantages such as ease of use or better performance for different dataset sizes and dimensionality [42]. In both cases, the AutoML methods were customized to maximize the Macro-F1 score (without the *green*-labeled posts). Each model was evaluated with 10-fold stratified cross-validation with five repeats inside of the training set. We trained the classifiers to predict the granular/fine-grained labels while evaluating the final output with the same macro-F1 score of the *amber*, *red*, and *crisis* categories.

We used the Tree-based Optimization Tool (TPOT) [43], which builds and selects machine learning pipelines using genetic programming. TPOT is built to generate pipelines that maximize classification accuracy while penalizing complex pipelines. Similarly, we used Auto-Sklearn to train and build classifiers using Bayesian optimization meta-learning and ensemble construction [44]. Given the high proportion of no risk labels in the datasets tested, we note that Auto-Sklearn contains a *Rebalancer* class for handling imbalanced class distributions. We primarily used default TPOT/Auto-Sklearn parameters with a population size of 200, a maximum evaluation time for a single pipeline of 5 min and total time as a stopping parameter, typically set to 2 days.

Mantel Tests

To compute the matrix of pairwise Euclidean distances between posts for each set of features, we used SciPy's distance matrix function [45]. This test allows quantification of the distances between posts across the various feature spaces. This is done in an unsupervised manner across the training and test posts. We used scikit-bio's mantel function with 999 permutations to perform the Mantel test on these distance matrices.

Emoji Visualization

To better understand the distribution of the 64 emoji features represented across the labeled posts, we aggregated the mean of an emoji feature across sentences in a post. Each of these aggregate features was then normalized to be between 0 and 1 to better compare features against each other. To obtain a measure of feature importance, we permuted each feature column and assessed the decrease in classification performance on the macro-F1 metric while using the best-performing pipeline derived from TPOT. For each emoji feature, we performed this procedure 10,000 times. Images of the emojis were obtained from EmojiOne (currently JoyPixels Inc) and converted to grayscale.

Availability

The CLPsych 2017 and UMD Reddit Suicidality datasets are available upon request from the original sources [28,29]. The code and instructions to fine-tune, train, and test a GPT-1 model on the CLPsych 2017 dataset is available online [46].

Results

Classification

To benchmark the performance of various text derived features for the automated classification of online forum posts, we ran both TPOT and Auto-Sklearn on the features generated from the post bodies. In Table 1, we report the average observed score

across the training folds, the final score on the held-out Reachout.com test set, and the external validation performance on Reddit data of the classifier trained only on Reachout.com data. In Figure 1, we present confusion matrices from 2 separate models trained with Auto-Sklearn to better demonstrate the predictions made across the imbalanced classes. Panel A shows the predictions of the VADER features, which resulted in a macro-F1 of 0.263. Panel B shows the predictions of the top-performing system with fine-tuned GPT features (a macro-F1 of 0.572).

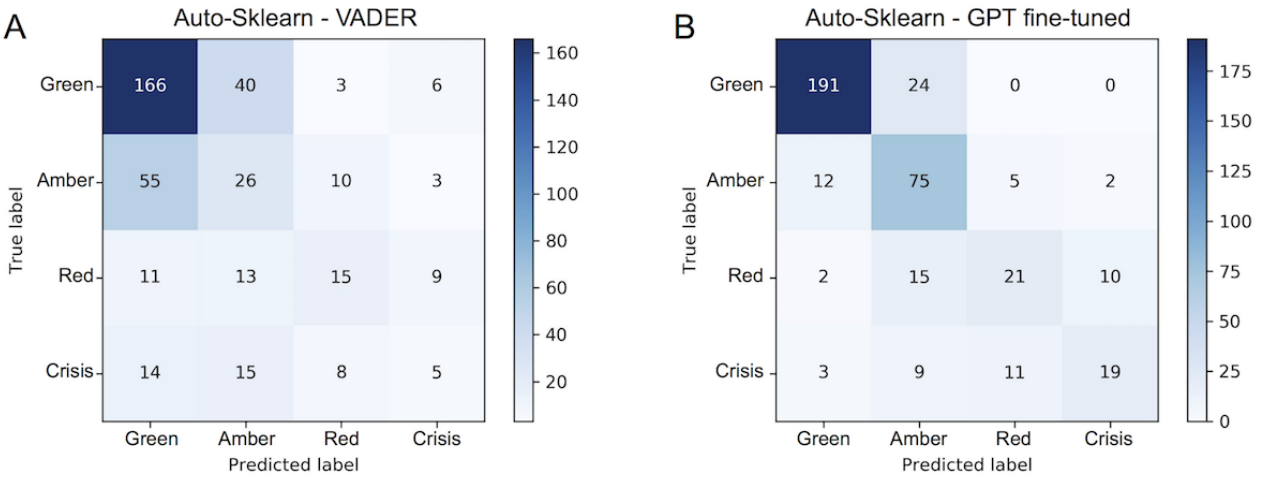
Table 1. Benchmarking by features, automated machine learning methods, and datasets with the macro-F1 metric.

Feature set	Feature count	Tree-based Optimization Tool			Auto-Sklearn		
		Train 10-fold, 5 times	Test	Reddit validation	Train 10-fold, 5 times	Test	Reddit validation
Empath (post)	195	0.280	0.253	0.385[a]	0.292	0.344	0.321
Linguistic Inquiry and Word Count	70	0.434	0.354	0.346[a]	0.433	0.380	0.315
Valence Aware Dictionary and sEntiment Reasoner (sentence)	12	0.363	0.263	0.356[a]	0.340	0.263	0.353[a]
Emoji 64	192	0.425	0.369	0.280	0.424	0.461	0.308
DeepMoji	6912	0.442	0.452	0.345[a]	0.391	0.437	0.351[a]
Universal Sentence Encoder	1536	0.457	0.446	0.300	0.484	0.479	0.236
GPT[b] default	2304	0.373	0.334	0.344[a]	0.396	0.383	0.402[a]
GPT fine-tuned	2304	0.510	0.559	0.320	0.492	0.572	0.324

[a]Reddit validation performance better than chance.

[b]GPT: Generative Pretrained Transformer.

Figure 1. Confusion matrices for 2 models trained with Auto-Sklearn. Each cell in the matrix provides the counts of posts that were labeled in the corresponding row and column axis that represent the predicted and true labels, respectively. Counts are colored from the highest cell (blue) to the lowest (white). The top-left to bottom-right diagonal cells count correctly predicted posts. Panel A trained with Valence Aware Dictionary and sEntiment Reasoner (VADER) features. Panel B trained with features from a fine-tuned Generative Pretrained Transformer (GPT) language model.



We noted that the average macro-F1 obtained during training was a fairly reliable predictor of the score on the held-out test set. Auto-Sklearn performed better on average than TPOT (mean test macro-F1 of 0.414 versus 0.379, respectively). We also observed the trend that features extracted from pretrained models perform better in general (average Auto-Sklearn test macro-F1 of 0.329 versus 0.466). However, the features extracted from the default GPT model (without any additional fine-tuning) were the worst performing of those obtained from neural models, whereas the GPT model that was fine-tuned on the unlabeled posts performed best across all experiments. The Universal Sentence Encoder and fine-tuned GPT features exceeded the highest macro-F1 score reached in the 2017 CLPsych-shared task when a classifier was learned with Auto-Sklearn (0.467; submission by Xianyi Xia and Dexi Liu). Upon inspection, the Auto-Sklearn–generated classifier for the GPT fine-tuned features was a complex ensemble of pipelines with multiple preprocessing steps and random forest classifiers. The TPOT-generated classification pipeline first selects features using the analysis of variance *F* value, then binarizes the values for classification with a K-nearest neighbor classifier (k=21; Euclidean distance). In contrast, the classifiers generated for

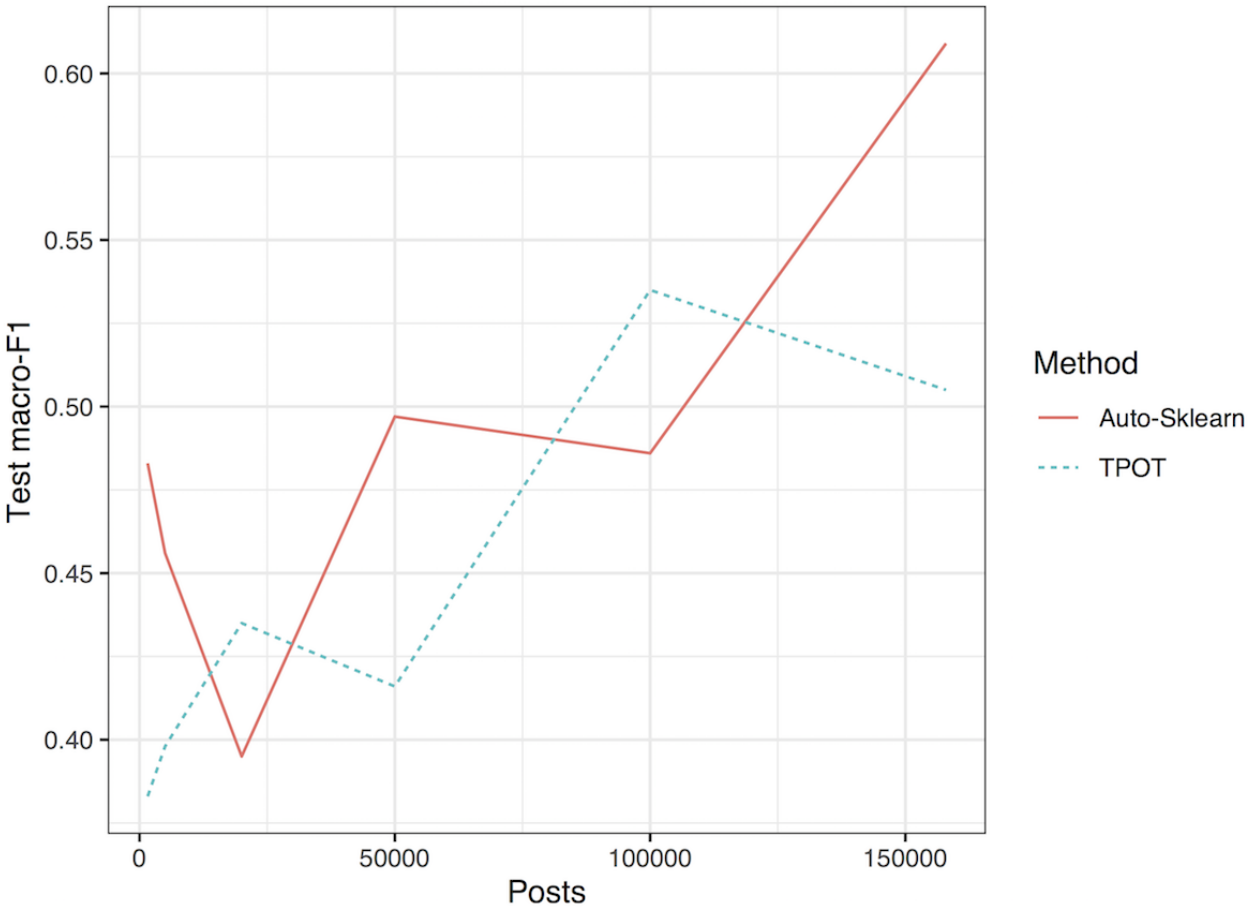
the Universal Sentence Encoder features are a linear support vector machine (TPOT) and ensembles of linear discriminant analysis classifiers (Auto-Sklearn).

To better understand the low Reddit validation scores, we calculated a random baseline. Although it is random, this does use information about the class distributions. We marked Reddit validation performance as better than chance in Table 1 with an [a]. Only classifiers learned from the VADER, DeepMoji, and default GPT features had macro-F1 scores above the threshold for both the TPOT and Auto-Sklearn learned classifiers. Unlike the CLPsych 2017 score that does not include the *green* or no risk labels, we used macro-F1 from all classes in the Reddit validation tests (corresponding to the CLPsych 2019 primary metric). When using the macro-F1 score that excluded the *no risk* class in the Reddit validation, none of the classifiers outperformed random runs at the same threshold. This is because of the classifiers having a good performance on the *no risk* or *green* labels and not the 3 remaining labels.

To better assess the variability of our best-performing system (Auto-Sklearn trained with features generated from the fine-tuned GPT model), we reran the Auto-Sklearn training and testing process 20 times. For each run, Auto-Sklearn was allotted 24 hours of compute time. Across those 20 systems, the average macro-F1 score on the held-out test set was 0.5293 (SD 0.0348). Of those 20 systems, the best- and worst-performing systems had a final test score of 0.6156 and 0.4594, respectively. Importantly, despite the variability and less compute time, the average macro-F1 score of these classifiers performed better than the scores obtained from different feature sets.

To determine the impact of the amount of data used for fine-tuning the GPT model on its effectiveness for feature extraction in the classification task, we fine-tuned models with increasing amounts of unlabeled posts before extracting post-level features to train a classifier (Figure 2). Although there is significant variability, there is a general trend of better performance when using models trained on a larger amount of unlabeled data.

Figure 2. A graph of macro-F1 test scores versus the number of posts used for Generative Pretrained Transformer-1 fine-tuning. Auto-Sklearn methods are marked with continuous red (Auto-Sklearn) and dashed blue (Tree-based Optimization Tool, TPOT) lines.



To compare the different representations or embeddings of the post contents, we used the Mantel test (Table 2). This compares the representations independently of their triage performance and suggests possible combinations for meta-classifiers. This test correlates the pairwise distances between posts in the benchmarked feature spaces, where a high correlation value between compared matrices indicates a significant overlap in the information they contain. Specifically, the Mantel test values range from −1 (perfect negative correlation) to 1 (perfect positive correlation), with zero representing no association between the pairs of posts in the feature spaces. Intriguingly, we observed the highest correlation between the Universal Sentence encoded features with those encoded by GPT. This is despite the comparison of aggregated DeepMoji encoded

features with aggregated 64-dimensional emoji encoding of DeepMoji, which we expected to have the strongest relationship. Similarly, comparisons between the default GPT and the fine-tuned version were slightly lower than correlations with the Universal Sentence Encoder. Although it is unclear, we presumed some of these differences may be due to the aggregation of sentence-level features into a post-level representation. None of the correlations with Empath features were significant, which probably reflects the sparsity of these features.

Table 2. Mantel correlations between the extracted feature sets.

Feature Set	VADER[a]	Empath	LIWC[b]	Universal Sentence	Emoji 64	DeepMoji	GPT[c] default	GPT fine-tuned
VADER	1.000	0.003	0.098	0.453	0.211	0.422	0.430	0.429
Empath	0.003	1.000	0.009	0.006	−0.005	−0.008	0.004	0.001
LIWC	0.098	0.009	1.000	0.148	0.403	0.507	0.267	0.253
Universal Sentence	0.453	0.006	0.148	1.000	0.193	0.509	0.823	0.823
Emoji 64	0.211	−0.005	0.403	0.193	1.000	0.523	0.302	0.335
DeepMoji	0.422	−0.008	0.507	0.509	0.523	1.000	0.632	0.631
GPT default	0.430	0.004	0.267	0.823	0.302	0.632	1.000	0.799
GPT fine-tuned	0.429	0.001	0.253	0.823	0.335	0.631	0.799	1.000

[a]VADER: Valence Aware Dictionary and sEntiment Reasoner.

[b]LIWC: Linguistic Inquiry and Word Count.

[c]GPT: Generative Pretrained Transformer.

System Interpretability

In Figure 3, we show the distribution of the mean emoji features for the top 10 most important features when using the mean emoji feature across sentences (64 total features). We noted that the interpretation and even visual representation of these emojis vary greatly, and these emojis were not used in the social media posts but were extracted by DeepMoji [38]. For example, the pistol emoji has been replaced by a ray gun or water gun in most platforms. From these distributions, it is clear that there is considerable variability across posts. This visualization also highlights the difficulty in discriminating the varying levels of risk when compared with the no risk posts. Of these top 10, 2 winking emojis are negatively correlated with risk, marking the importance of a positive sentiment. As expected, the negative emojis are more important, with the pistol, skull, and broken heart emoji ranked in the top 5.

To better understand judgments made by our trained classifier, we present predictions in Figure 4 on a set of composite quotes and their themes from a study of suicide notes [34]. For each quote, we presented the initial prediction (with the granular/fine-grained prediction in parentheses). Across the 10 quotes, 3 were classified as crisis, 4 as red, and 3 as amber. One of the amber classifications is under the "Hopelessness secondary to chronicity of illness and treatment" theme, further suggesting that our system may not recognize expressions of hopelessness.

All words were iteratively masked to indicate their effects on the predicted class (see Methods section). In Figure 4, words that affected predictions are color coded. The colored words are important for indicating severity as removing them makes the quotes appear less severe to our system. Examining these words suggests that negations affected severity (eg, "not," "can't"). In the quotes, negations seemed to indicate a perceived failure or not having done or achieved something the person felt they ought to. Expressions of hopelessness (ie, "no hope left") were also important in classifying quotes as severe by our system. Words reflecting an unwillingness or inability to continue were also important (ie, "I'm done," "I am too tired to") as were words indicating loneliness (ie, "being isolated"). In contrast, replacing a green word with an unknown word shifted the predicted class to a more severe category (eg, from *red* to *crisis*). On examining the nature of the green words (ie, "what," "after"), it was not clear why these words were important for lessening the severity of the quotes.

For 2 of the quotes predicted as red, no words were highlighted, suggesting that, in these instances, many words were key to the prediction. Overall, the quotes would all be flagged as requiring some level of moderator attention, and for the most part, the nature of words that were important in classifying the severity of quotes made conceptual sense.

Figure 3. Violin plot showing the distributions of the 10 most discriminative emoji features across labeled classes. The classes are according to label with crisis in gray. The y-axis is the predicted scores for each emoji that have been scaled to the 0-1 interval. The emojis across the y-axis are marked with their images and their official Unicode text labels. The emojis are ranked from the most to least important feature (left to right).

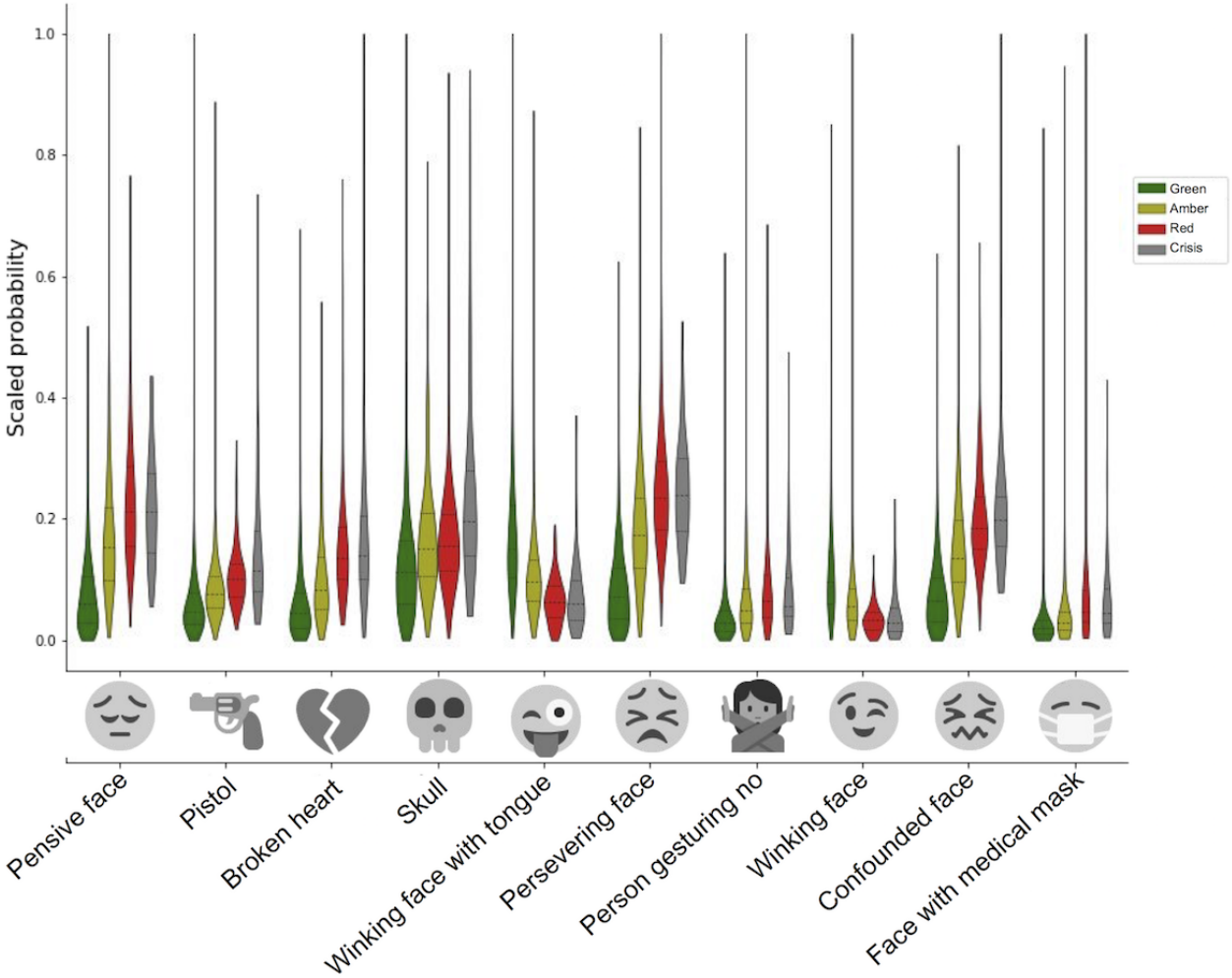


Figure 4. Predictions and highlights of suicide-related composite quotes from Furqan and colleagues. Words that changed predictions are color coded. Replacing a yellow or red word with an unknown word shifts the prediction to a less severe class by 1 or 2 levels, respectively, (ie, replacing a yellow word in text that is classified as crisis would change the prediction to red while a red word would change it to amber). In contrast, replacement of green words will result in more severe predictions.

Theme	Prediction	Composite quote
Suicide as exertion of personal autonomy	Red (currentAcuteDistress)	"I am the person who decides what I m going to do with my life. After being isolated and marginalized, I m done with it all. This is my life and my decision."
Exhaustion	Red (currentAcuteDistress)	"I am exhausted from trying to fix everything. No one understands, not even doctors or my family, and I keep trying to get help. I feel like I m a dead man walking for a long time. I ve been judged by society and have been made to feel ashamed. I m burned out and tired and trying to find some way to rest."
Treatment failure as personal failure	Crisis	"It does n t matter how many treatments I try , there is something so wrong with me. It s not the treatments, it s me."
Treatment failure as personal failure	Crisis	"Medication, therapy, counseling, alcohol I ve tried everything and I ca n t seem to figure out a way out of this, I ca n t crack the code."
Hopelessness secondary to chronicity of illness and treatment	Amber (followupOk)	"After trying multiple treatments , without curing my mental illness, I have realized that a solution other than death just does not exist."
Conflict between Self and Illness	Crisis	"I have fought against my thoughts, depression and alcohol constantly. I am too tired to keep going"
Illness as biological	Amber (followupOk)	"There is a problem in my brain. I think the chemistry is all wrong. No one could have fixed that."
Low levels of perceived agency	Amber (followupOk)	"Forgive me. It s not my fault, it s a disease, I hope you can understand that."
Low levels of perceived agency	Red (followupWorse)	"I sank into a deep depression. It was the depression that ruined my relationship. I want my family to know that."
High levels of perceived agency	Red (currentAcuteDistress)	"I had so much in my life family, friends , career but I let the disease, addiction, and my own personality take over me and it ruined everything in my life. Looking back, there were times that I should have changed the course of my life but I did n t and now there is no hope left."

Discussion

We have shown that there are highly informative signals in the text body alone of posts from the Reachout.com forum. More specifically, we identified a transfer learning approach as particularly useful for extracting features from raw social media text. In combination with the training of classifiers using AutoML methods, we showed that these representations of the post content can improve triage performance without considering the context or metadata of the posts. These methods take advantage of the large amount of unlabeled free text that is often available to diminish the need for labeled examples. We also showed that these methods can generalize to new users on a

support forum, for which there would not be preceding posts to provide context on their mental states. By combining the pretrained language models with AutoML, we were able to achieve state-of-the-art macro-F1 on the CLPsych 2017 shared task. Our content-only approach could be complemented by previous work, which used hand-engineered features to account for contextual information, such as a user's post history or the thread context of posts [26,47]. Future developments could also include multiple types of media (eg, text, photos, videos) that are often present on social media to better assess the subtleties of users' interactions [48].

Our current approach follows methods outlined by Radford et al [20] to fine-tune the language model that was previously pretrained on a large corpus of books. This fine-tuning step allows the model to learn the characteristics of the Reachout.com text. We show that increasing the amounts of in-domain unlabeled data for fine-tuning improves classification performance and has yet to reach a plateau. Further work will be instrumental in defining when and how to fine-tune pretrained language models better [49]. For tasks with limited data availability, the ability to adapt and fine-tune a model on multiple intermediate tasks could be a particularly worthwhile approach, as demonstrated by the Universal Sentence Encoder and others [39,50]. However, it is unclear how these large language models can retain and accumulate knowledge across tasks and datasets. Notably, it has been reported that these large pretrained language models are difficult to fine-tune and that many random restarts may be required to achieve optimal performance [51,52].

We compared the use of AutoML tools, such as Auto-Sklearn and TPOT, to generate classification pipelines with a variety of features extracted from free text. We also identified them as sources of variability in the final scores of our system. When developing our top-performing systems with features extracted from a fine-tuned GPT and using Auto-Sklearn on 20 trials, we obtained macroaverage F1 scores ranging from 0.6156 to 0.4594. In part, this is because of the small size of the dataset and the weighted focus of the macroaverage F1 metric toward the *crisis* class with relatively fewer instances. Further experiments, although computationally intensive, could help distinguish the amount of variability that is inherent in the language model fine-tuning process.

There are a variety of limitations, depending on the use of the approaches we benchmarked. Further experiments would be needed to determine if Reachout.com moderator responsiveness improves when more accurate classifiers are used. The present system performance cannot be extrapolated too far into the future because of changes in the population of users on the forum, shifting topics discussed or variations in language used. Furthermore, it is important to note that any implemented system would require ongoing performance monitoring.

To further understand how our trained models would perform in a new context, we assessed performance on an independently collected dataset and composite quotes that were derived from suicide notes. All composite quotes were flagged as requiring moderator attention. Our classifiers generalize to some degree on the UMD Reddit Suicidality Dataset, which approximates the task outlined for Reachout.com. We noted that the Reddit user base is not specific to Australia, is not targeted explicitly to youth, and may have substantially different topics of discussion than Reachout.com. This performance is primarily driven by good accuracy on the *no risk* or *green* class. We observed that the features derived from the fine-tuned GPT model perform worse than those from the default GPT model, indicating that this model might be specific to unique features of Reachout.com. Future studies could determine whether multiple rounds of fine-tuning on different datasets increase accuracy.

We manually reviewed the errors made by the best-performing system (Auto-Sklearn classifier with the GPT fine-tuned features). The most worrisome prediction errors occur when the classifier mistakes a crisis post for one of lesser importance, which could potentially delay a moderator response. When posts were not classified as crisis posts (but should have been), this was often due to vague language referring to self-harm or suicide (eg, "time's up," "get something/do it," "to end it," "making the pain worse"). Sometimes, forum users deliberately referred to self-harm or suicide with nonstandard variations, such as "SH" or "X" (eg, "attempt X," "do X"). Future work could be instructive in determining whether these words are associated with higher levels of distress/crisis relative to the words they are meant to replace. Alternatively, custom lexicons might be developed to capture instances of self-harm or suicide represented by vague language or nonstandard variations.

In some failure cases (ie, posts that should be classified as being of higher risk than they were), the classifier did not notice expressions of hopelessness, which may cue the imminence of risk. Other prominent failure cases were instances when the classifier did not notice a poster's dissatisfaction with mental health services that provide real-time help (eg, suicide call-back services and crisis helplines, etc). According to the labeling scheme, these posts should be classified as red. However, this dissatisfaction was often conveyed in diverse and highly contextualized ways, likely making it difficult for the system to identify. There were also posts that did not indicate imminent risk but described sensitive topics such as feeling lonely or losing a parent. These were often misclassified as green (when they should have been amber), possibly because they also contained positive language, or the sensitivity of the topic was difficult for the system to grasp.

In some of these failure cases, it may have been useful to take into account the previous post; eg, when the post in question is short or vague, the system may classify the level of risk more accurately if the previous post expresses a high level of concern about the poster or tries to convince the poster to seek immediate help.

Neural networks can build complex representations of their input features, and it can be difficult to interpret how these representations are used in the classification process. In a deeper analysis of DeepMoji features, we identified the most important emoji for classification and found that the emotional features follow a linear arrangement of expression at the class level corresponding to label severity. We also used input masking to iteratively highlight the contributions of individual words to the

final classification. Such highlighting and pictorial/emoji visualizations could speed moderator review of posts. Ultimately, we believe the further development of methods to improve model interpretability will be essential in facilitating the work of mental health professionals in Web-based contexts.

In conclusion, we showed that transfer learning combined with AutoML provides state-of-the-art performance on the CLPsych 2017 triage task. Specifically, we found that an AutoML classifier trained on features from a fine-tuned GPT language model was the most accurate. We suggest this automated transfer learning approach as the first step to those building natural language processing systems for mental health because of the ease of implementation. Although such systems lack interpretability, we showed that emoji-based visualizations and masking can aid explainability.

Acknowledgments

The Centre for Addiction and Mental Health (CAMH) Specialized Computing Cluster, which is funded by the Canada Foundation for Innovation and the CAMH Research Hospital Fund, was used to perform this research. The authors thank the Nvidia Corporation for the Titan Xp GPU that was used for this research. The authors acknowledge the assistance of the American Association of Suicidology in making the University of Maryland Reddit Suicidality Dataset available. The authors also thank the 3 anonymous reviewers for their helpful suggestions and comments. This study was supported by the CAMH Foundation and a National Science and Engineering Research Council of Canada Discovery Grant to LF.

Conflicts of Interest

LF owns shares in Alphabet Inc, which is the parent company of Google, the developer of the freely available Universal Sentence Encoder, which was compared with other methods.

Multimedia Appendix 1
Table including fine-grained annotations of Reachout posts.
[XLSX File (Microsoft Excel File), 4 KB - jmir_v22i5e15371_app1.xlsx]

References

1. Kessler RC, Wang PS. The descriptive epidemiology of commonly occurring mental disorders in the United States. Annu Rev Public Health 2008;29:115-129. [doi: 10.1146/annurev.publhealth.29.020907.090847] [Medline: 18348707]
2. Plana-Ripoll O, Pedersen CB, Holtz Y, Benros ME, Dalsgaard S, de Jonge P, et al. Exploring comorbidity within mental disorders among a Danish national population. JAMA Psychiatry 2019 Mar 1;76(3):259-270 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.3658] [Medline: 30649197]
3. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):617-627 [FREE Full text] [doi: 10.1001/archpsyc.62.6.617] [Medline: 15939839]
4. Kessler RC, Amminger GP, Aguilar-Gaxiola S, Alonso J, Lee S, Ustün TB. Age of onset of mental disorders: a review of recent literature. Curr Opin Psychiatry 2007 Jul;20(4):359-364 [FREE Full text] [doi: 10.1097/YCO.0b013e32816ebc8c] [Medline: 17551351]
5. Nock MK, Hwang I, Sampson NA, Kessler RC. Mental disorders, comorbidity and suicidal behavior: results from the National Comorbidity Survey Replication. Mol Psychiatry 2010 Aug;15(8):868-876 [FREE Full text] [doi: 10.1038/mp.2009.29] [Medline: 19337207]
6. Franklin JC, Ribeiro JD, Fox KR, Bentley KH, Kleiman EM, Huang X, et al. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. Psychol Bull 2017 Feb;143(2):187-232. [doi: 10.1037/bul0000084] [Medline: 27841450]
7. Naslund JA, Aschbrenner KA, Marsch LA, Bartels SJ. The future of mental health care: peer-to-peer support and social media. Epidemiol Psychiatr Sci 2016 Apr;25(2):113-122 [FREE Full text] [doi: 10.1017/S2045796015001067] [Medline: 26744309]
8. Davidson L, Chinman M, Sells D, Rowe M. Peer support among adults with serious mental illness: a report from the field. Schizophr Bull 2006 Jul;32(3):443-450 [FREE Full text] [doi: 10.1093/schbul/sbj043] [Medline: 16461576]
9. Kaplan K, Salzer MS, Solomon P, Brusilovskiy E, Cousounis P. Internet peer support for individuals with psychiatric disabilities: a randomized controlled trial. Soc Sci Med 2011 Jan;72(1):54-62. [doi: 10.1016/j.socscimed.2010.09.037] [Medline: 21112682]
10. Ridout B, Campbell A. The use of social networking sites in mental health interventions for young people: systematic review. J Med Internet Res 2018 Dec 18;20(12):e12244 [FREE Full text] [doi: 10.2196/12244] [Medline: 30563811]
11. de Choudhury M, Counts S, Horvitz E. Predicting Postpartum Changes in Emotion and Behavior via Social Media. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2013 Presented at: CHI'13; April 27 - May 2, 2013; Paris, France p. 3267-3276. [doi: 10.1145/2470654.2466447]

12. Kornfield R, Sarma PK, Shah DV, McTavish F, Landucci G, Pe-Romashko K, et al. Detecting recovery problems just in time: application of automated linguistic analysis and supervised machine learning to an online substance abuse forum. J Med Internet Res 2018 Jun 12;20(6):e10136 [FREE Full text] [doi: 10.2196/10136] [Medline: 29895517]
13. Milne DN, McCabe KL, Calvo RA. Improving moderator responsiveness in online peer support through automated triage. J Med Internet Res 2019 Apr 26;21(4):e11410 [FREE Full text] [doi: 10.2196/11410] [Medline: 31025945]
14. Cohan A, Desmet B, Yates A, Soldaini L, MacAvaney S, MacAvaney S, et al. arXiv e-Print archive. 2018. SMHD: A Large-Scale Resource for Exploring Online Language Usage for Multiple Mental Health Conditions URL: http://arxiv.org/abs/1806.05258 [accessed 2020-02-28]
15. Gkotsis G, Oellrich A, Hubbard TJ, Dobson RJ, Liakata M, Velupillai S, et al. The Language of Mental Health Problems in Social Media. In: Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology. 2016 Presented at: CLPsych'16; June 16, 2016; San Diego, California p. 63-73 URL: https://www.aclweb.org/anthology/W16-0307 [doi: 10.18653/v1/W16-0307]
16. Yates A, Cohan A, Goharian N. Depression and Self-Harm Risk Assessment in Online Forums. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017 Presented at: EMNLP'17; September 7–11, 2017; Copenhagen, Denmark URL: http://arxiv.org/abs/1709.01848 [doi: 10.18653/v1/d17-1322]
17. Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. J Lang Soc Psychol 2010;29(1):24-54. [doi: 10.1177/0261927X09351676]
18. Edwards T, Holtzman NS. A meta-analysis of correlations between depression and first person singular pronoun use. J Res Pers 2017 Jun;68:63-68 [FREE Full text] [doi: 10.1016/j.jrp.2017.02.005]
19. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed Representations of Words and Phrases and their Compositionality. In: Proceedings of the Advances in Neural Information Processing Systems 26. 2013 Presented at: NIPS'13; Dec 5-10, 2013; Lake Tahoe, Nevada URL: http://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf
20. Radford A, Narasimhan K, Salimans T, Sutskever I. Computer Science at UBC. 2018. Improving Language Understandingby Generative Pre-Training URL: https://www.cs.ubc.ca/~amuham01/LING530/papers/radford2018improving.pdf [accessed 2020-02-20]
21. Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans Knowl Data Eng 2010 Oct;22(10):1345-1359. [doi: 10.1109/TKDE.2009.191]
22. Peters ME, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, et al. arXiv e-Print archive. Deep Contextualized Word Representations URL: http://arxiv.org/abs/1802.05365 [accessed 2020-02-20]
23. Howard J, Ruder S. arXiv e-Print archive. 2018. Universal Language Model Fine-Tuning for Text Classification URL: http://arxiv.org/abs/1801.06146 [accessed 2020-02-20]
24. Milne DN, Pink G, Hachey B, Calvo RA. CLPsych 2016 Shared Task: Triaging Content in Online Peer-Support Forums. In: Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology. 2016 Presented at: CLPsych'16; June 16, 2016; San Diego URL: https://www.aclweb.org/anthology/W16-0312
25. Kessler RC, Berglund P, Borges G, Nock M, Wang PS. Trends in suicide ideation, plans, gestures, and attempts in the United States, 1990-1992 to 2001-2003. J Am Med Assoc 2005 May 25;293(20):2487-2495. [doi: 10.1001/jama.293.20.2487] [Medline: 15914749]
26. Altszyler E, Berenstein AJ, Milne D, Calvo RA, Slezak DF. Using Contextual Information for Automatic Triage of Posts in a Peer-Support Forum. In: Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic. 2018 Presented at: CLPsych NAACL'18; June 2018; New Orleans, LA URL: https://www.aclweb.org/anthology/papers/W/W18/W18-0606/
27. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. Fam Med 2005 May;37(5):360-363 [FREE Full text] [Medline: 15883903]
28. Zirikly A, Resnik P, Uzuner O, Hollingshead K. CLPsych 2019 Shared Task: Predicting the Degree of Suicide Risk in Reddit Posts. In: Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology.: Association for Computational Linguistics; 2019 Presented at: CLPsych NAACL'19; June 6, 2019; Minneapolis, Minnesota.
29. Shing HC, Nair S, Zirikly A, Friedenberg M, Daumé H, Resnik P. Expert, Crowdsourced, and Machine Assessment of Suicide Risk via Online Postings. In: Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic.: Association for Computational Linguistics; 2018 Presented at: CLPsych NAACL'18; June 5 - 6, 2018; New Orleans, LA p. 25-36. [doi: 10.18653/v1/W18-0603]
30. Corbitt-Hall DJ, Gauthier JM, Davis MT, Witte TK. College students' responses to suicidal content on social networking sites: an examination using a simulated Facebook newsfeed. Suicide Life Threat Behav 2016 Oct;46(5):609-624. [doi: 10.1111/sltb.12241] [Medline: 26999294]
31. Krippendorff K. Reliability in content analysis. Human Comm Res 2004;30(3):411-433 [FREE Full text] [doi: 10.1111/j.1468-2958.2004.tb00738.x]
32. Dawid AP, Skene AM. Maximum likelihood estimation of observer error-rates using the EM Algorithm. Appl Stat 1979;28(1):20-28 [FREE Full text] [doi: 10.2307/2346806]

33. Passonneau RJ, Carpenter B. The benefits of a model of annotation. Trans Assoc Comput Linguist 2014;2:311-326 [FREE Full text] [doi: 10.1162/tacl_a_00185]
34. Furqan Z, Sinyor M, Schaffer A, Kurdyak P, Zaheer J. 'I Can't Crack the Code': what suicide notes teach us about experiences with mental illness and mental health care. Can J Psychiatry 2019 Feb;64(2):98-106 [FREE Full text] [doi: 10.1177/0706743718787795] [Medline: 30009639]
35. Hutto CJ, Gilbert E. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. In: Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media. 2014 Presented at: AAAI'14; July 27–31, 2014; Ann Arbor, Michigan URL: https://pdfs.semanticscholar.org/a6e4/a2532510369b8f55c68f049ff11a892fefeb.pdf?_ga=2.235950529.1395435436.1582877966-1679671381.1567599385
36. Pennebaker JW, Booth RJ, Francis ME. Texas Tech University Departments. Linguistic Inquiry and Word Count: LIWC2007 URL: http://www.depts.ttu.edu/psy/lusi/files/LIWCmanual.pdf [accessed 2020-02-20]
37. Fast E, Chen B, Berstein MS. Empath: Understanding Topic Signals in Large-Scale Text. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016 Presented at: CHI'16; May 7 - 12, 2016; San Jose, California p. 4647-4657. [doi: 10.1145/2858036.2858535]
38. Felbo B, Mislove A, Søgaard A, Rahwan I, Lehmann S. arXiv e-Print archive.: arXiv; 2017. Using Millions of Emoji Occurrences to Learn Any-domain Representations for Detecting Sentiment, Emotion and Sarcasm URL: http://arxiv.org/abs/1708.00524 [accessed 2020-02-20]
39. Cer D, Yang Y, Kong SY, Hua N, Limtiaco N, St John R, et al. arXiv e-Print archive.: arXiv; 2018. Universal Sentence Encoder URL: http://arxiv.org/abs/1803.11175 [accessed 2020-02-20]
40. May M, Townsend BL, Matthew B. GitHub.: Indico Data Solutions Finetune: Scikit-Learn Style Model Finetuning for NLP URL: https://github.com/IndicoDataSolutions/finetune [accessed 2020-02-20]
41. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res 2011;12:2825-2830 [FREE Full text]
42. Truong A, Walters A, Goodsitt J, Hines K, Bruss CB, Farivar R. arXiv e-Print archive.: arXiv; 2019. Towards Automated Machine Learning: Evaluation and Comparison of AutoML Approaches and Tools URL: http://arxiv.org/abs/1908.05557 [accessed 2020-02-20]
43. Olson RS, Urbanowicz RJ, Andrews PC, Lavender NA, Kidd LC, Moore JH. Automating biomedical data science through tree-based pipeline optimization. In: Squillero G, Burelli P, editors. Applications of Evolutionary Computation. Cham: Springer; 2016:123-137.
44. Feurer M, Klein A, Eggensperger K, Springenberg J, Blum M, Hutter F. Efficient and Robust Automated Machine Learning. In: Proceedings of the Advances in Neural Information Processing Systems 28. 2015 Presented at: NIPS'15; December 7-12, 2015; Montreal URL: http://papers.nips.cc/paper/5872-efficient-and-robust-automated-machine-learning.pdf
45. Jones E, Oliphant T, Peterson P. SciPy: Open source scientific tools for Python. ScienceOpen 2001:- [FREE Full text]
46. GitHub. URL: https://github.com/derekhoward/Reachout_triage [accessed 2020-03-01]
47. Amir S, Coppersmith G, Carvalho P, Silva MJ, Wallace BC. arXiv e-Print archive.: arXiv; 2017. Quantifying Mental Health from Social Media with Neural User Embeddings URL: http://arxiv.org/abs/1705.00335 [accessed 2020-02-20]
48. Chancellor S, Kalantidis Y, Pater JA, de Choudhury M, Shamma DA. Multimodal Classification of Moderated Online Pro-Eating Disorder Content. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI'17; May 6 - 11, 2017; Denver Colorado USA p. 3213-3226. [doi: 10.1145/3025453.3025985]
49. Peters ME, Ruder S, Smith NA. arXiv e-Print archive.: arXiv; 2019. To Tune or Not to Tune? Adapting Pretrained Representations to Diverse Tasks URL: http://arxiv.org/abs/1903.05987 [accessed 2020-02-20]
50. Yogatama D, d'Autume CM, Connor J, Kocisky T, Chrzanowski M, Kong L, et al. arXiv e-Print archive.: arXiv Learning and Evaluating General Linguistic Intelligence URL: http://arxiv.org/abs/1901.11373 [accessed 2020-02-20]
51. Devlin J, Chang MW, Lee K, Toutanova K. arXiv e-Print archive.: arXiv; 2018. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding URL: http://arxiv.org/abs/1810.04805 [accessed 2020-02-20]
52. Phang J, Févry T, Bowman SR. arXiv e-Print archive.: arXiv; 2018. Sentence Encoders on STILTs: Supplementary Training on Intermediate Labeled-Data Tasks URL: http://arxiv.org/abs/1811.01088 [accessed 2020-02-20]

Abbreviations

AutoML: automated machine learning
CAMH: Centre for Addiction and Mental Health
CLPsych: Computational Linguistics and Clinical Psychology
GPT: Generative Pretrained Transformer
LIWC: Linguistic Inquiry and Word Count
TPOT: Tree-based Optimization Tool
VADER: Valence Aware Dictionary and sEntiment Reasoner

Edited by G Eysenbach; submitted 04.07.19; peer-reviewed by A Jaroszewski, E Kleiman, N Miyoshi; comments to author 21.10.19; revised version received 13.12.19; accepted 28.01.20; published 13.05.20.

Please cite as:
Howard D, Maslej MM, Lee J, Ritchie J, Woollard G, French L
Transfer Learning for Risk Classification of Social Media Posts: Model Evaluation Study
J Med Internet Res 2020;22(5):e15371
URL: https://www.jmir.org/2020/5/e15371
doi:10.2196/15371
PMID:32401222

Original Paper

Privacy Protection in Online Health Communities: Natural Experimental Empirical Study

Yuanyuan Dang[1], MA; Shanshan Guo[1], PhD; Xitong Guo[1], PhD; Doug Vogel[1], PhD

School of Management, Harbin Institute of Technology, Harbin, China

Corresponding Author:
Shanshan Guo, PhD
School of Management
Harbin Institute of Technology
92 West Dazhi Street
Nan Gang District
Harbin, 15001
China
Phone: 86 18845594033
Fax: 86 045186414024
Email: hit_guoshanshan@163.com

Abstract

Background: An online health community (OHC) is a novel sharing channel through which doctors share professional health care knowledge with patients. While doctors have the authority to protect their patients' privacy in OHCs, we have limited information on how doctors' privacy protection choices affect their professional health care knowledge sharing with patients.

Objective: We examined the relationship between privacy protection and professional health care knowledge sharing in OHCs. Specifically, we examined the effects of privacy protection settings in an OHC on doctors' interactive professional health care knowledge sharing and searching professional health care knowledge sharing (two dimensions of professional health care knowledge sharing). Moreover, we explored how such effects differ across different levels of disease stigma.

Methods: We collected the monthly panel data of 19,456 doctors from Good Doctor, one of the largest OHCs in China, from January 2008 to April 2016. A natural experimental empirical study with difference-in-difference analysis was conducted to test our hypotheses. The time fixed effect and the individual fixed effect were both considered to better identify the effects of a privacy protection setting on professional health care knowledge sharing. Additionally, a cross-sectional analysis was performed for a robust check.

Results: The results indicate that the privacy protection setting has a significant positive effect on interactive professional health care knowledge sharing (β=.123, *P*<.001). However, the privacy protection setting has a significant negative effect on searching professional health care knowledge sharing (β=–.225, *P*=.05). Moreover, we found that high disease stigma positively impacts the effect of privacy protection on interactive professional health care knowledge sharing (coefficients are in the same valence) and negatively impacts the effects of privacy protection on searching professional health care knowledge sharing (coefficients are in the reverse valence).

Conclusions: Privacy protection has a bilateral effect on professional health care knowledge sharing (ie, a positive effect on interactive professional health care knowledge sharing and a negative effect on searching professional health care knowledge sharing). Such bilateral switches of professional health care knowledge sharing call for a balanced state in conjunction with practical implications. This research also identifies a moderate effect of disease stigma on privacy protection settings and professional health care knowledge sharing.

(J Med Internet Res 2020;22(5):e16246) doi:10.2196/16246

KEYWORDS

online health community; privacy protection; professional health care knowledge

Introduction

Online Health Communities as Platforms for Sharing Professional Health Care Knowledge

The internet has dynamically invaded, affected, and even changed many traditional industries, such as business (electronic commerce [e-commerce]), finance (internet finance), social (virtual communities), and health care (telemedicine). Compared with the traditional offline mode, the capabilities of the internet (ie, its real-time communication across geographical boundaries, wide and swift spread, easy access, and rapid generation of users) have proven successful in the online mode. Recently, studies in both practice and research have focused on another internet capability: sharing, especially in the sharing economy. For instance, travel apps help in sharing transportation tools (eg, Uber), cloud computing apps help in sharing idle computing capacities (eg, Alibaba Cloud computing), and accommodation apps help in sharing lodgings (eg, Airbnb). In addition, regarding the health care industry, the online health community (OHC) (a kind of telemedicine-providing, doctor-patient means of communication using information and communication technology [1]) provides an online platform for sharing professional health care knowledge.

Similarly, with transportation tools, computing capacities, and lodgings, professional health care knowledge is a scarce and valuable resource. Having abundant and correct professional health care knowledge is conducive to enhancing coping and self-efficacy [2], affecting health-related decisions and the behavior of OHC users and their friends and family [3], enabling better management of chronic health conditions [4], and fueling discussions with health care providers [5]; health care, in general, is thereby maintained. However, most professional health care knowledge networks are owned by a small number of health care professionals. In traditional offline health care activities, professional health care knowledge is mainly shared through face-to-face treatment. In particular, it makes for inefficient sharing in a one-to-one context.

According to the definition of knowledge sharing [6,7], professional health care knowledge sharing is defined as medical professionals (eg, doctors, nurses, and pharmacists) sharing their professional health care knowledge with other health care stakeholders in the form of various activities (eg, treatment, consultations, and published papers) [8]. Fortunately, the OHC provides more possibilities for professional health care knowledge sharing, including the seeking of advice online and by telephone consultation [9,10]; for connecting with other patients with similar experiences [11,12]; and for the understanding of professional diagnoses by reading doctors' published papers. Moreover, the OHC supplies to the public the online interaction records between doctors and patients to help patients with similar ailments [13,14]. From the perspective of doctors, the professional health care knowledge sharing between doctors and patients can be divided into interactive professional health care knowledge sharing and searching professional health care knowledge sharing. Interactive professional health care knowledge sharing refers to the professional health care knowledge sharing process actively carried out by doctors through interaction with patients, such as online consultations and telephone consultations in OHCs. Searching professional health care knowledge sharing, on the other hand, refers to a process where there are no doctor-patient interactions; patients search, learn, and acquire doctors' professional health care knowledge autonomously, such as by searching doctors' published articles and doctor-patient interaction records (in Chinese OHCs, such doctor-patient interaction records, which are anonymously processed, are on display to all visitors by default, like those in the forums). Professional health care knowledge sharing, based on the internet, is being adopted by a growing number of users [15].

Privacy Protection and Professional Health Care Knowledge Sharing in OHCs

Unlike transportation tools, computing capacities, and lodgings, professional health care knowledge sharing is more inclined to be privacy-sensitive. It is more sensitive than other forms of information on online platforms, such as demographic profiles, lifestyle interests, or purchase history information [16]. Just as in an offline context, doctors make a diagnosis according to a patient's conditions; therefore, professionals may ask their patients to provide more personal data in exchange for the utility of the services. To obtain better treatment, manage their conditions, or improve their overall health, patients have to provide their personal information (such as gender, lifestyle, occupation, disease severity, and medical examination results), which may contain more sensitive personal information [17]. Moreover, professionals also post sensitive data about their patients, which is made visible to the public, so that advice on clinical situations or practice management may be shared with other patients or professionals [18,19].

OHCs are opening the door for inappropriate access, misuse, and disclosure of personal privacy [20]. Once such data are handled improperly, patients will be faced with privacy invasion [21], such as negative social stigma, as well as straining of family ties, losing control of medical information, and suffering harassment from commercial advertising [16]. These concerns can even cause individuals to avoid the health care services of OHCs, and thus limit the role of an OHC in professional health care knowledge sharing. Consequently, some OHCs have provided privacy protection mechanisms, such as improving data storage security, building a safe environment, and providing anonymity for patients. In addition, an OHC can enable doctors to activate their patients' privacy protection settings by making the consultation process invisible.

Objectives and Research Questions

Several studies (further previous research can be seen in Multimedia Appendix 1) have investigated professional health care knowledge sharing in OHCs among patients or doctors, such as the patient motivations for health care knowledge sharing on Wikipedia [22] and the development of doctors' knowledge sharing in virtual communities [23,24]. However, a dearth of studies concentrating on professional health care knowledge sharing between doctors and patients remains [25]. Moreover, although studies have explored the positive impact of privacy protection mechanisms on patients' OHC participation intentions and continuous use intentions [16,21],

privacy protection is also a double-edged sword [26] for professional health care knowledge sharing. Doctors' privacy protection settings in OHCs regarding patients' personal health care information will increase patient trust and promote knowledge sharing. However, it also blocks other patients' access to professional health care knowledge in their search for consultation records. Therefore, it is necessary to explore the influence of doctors' choice of privacy protection settings in OHCs on professional health care knowledge sharing between doctors and patients. To fill this research gap, it is essential to understand the nature of professional health care knowledge sharing in OHCs. Hence, we conducted an empirical analysis of a large OHC to explore the effects of privacy protection settings on interactive and searching professional health care knowledge sharing. In further extending our model, we investigated how professional health care knowledge is differentially shared with patients suffering from different levels of disease stigma. For instance, patients with different diseases (eg, influenza versus HIV) may have different professional health care knowledge sharing trajectories from the influence of the associated stigma.

The Enhancing Power of Privacy Protection on Interactive Professional Health Care Knowledge Sharing

In OHCs, interactive forms of professional health care knowledge sharing are mainly conducted through online doctor-patient consultations. In other words, an online consultation thread can be started for a doctor-and-patient pair, and each can interact with the other through text descriptions and the uploading of test and imaging results. Doctors can make judgments and recommendations based on their patients' descriptions of their symptoms to generate professional health care knowledge sharing in their interactions with patients. Without privacy protection, the entire online consultation process described above would be rendered publicly available to all internet users. However, as we mentioned previously, this may be harmful in terms of patients' privacy disclosure. The privacy protection mechanism of an OHC gives doctors the right to set up a transparent consultation process, so as to protect the privacy of their patients.

Previous studies indicated that a privacy policy can improve user trust and loyalty in health care providers, thus influencing health information exchange behaviors in users [27]. In the interactive professional health care knowledge sharing process of an OHC, patients may increase their trust in doctors when they are protected by doctors' privacy protection settings, thus feeling encouraged to share more personal health care information during communication [28]. The feeling of being protected will also increase patients' evaluations of doctors and render them willing to communicate with doctors more frequently, thus increasing the shared professional health care knowledge in the interactive process. Further, doctors' reasonable privacy protection settings will attract more patients to consult online for professional health care knowledge sharing. Hence, by increasing patient interaction intentions, consultation times, and the number of patients, doctors' privacy protection settings can affect the professional health care knowledge sharing frequency of the doctor-patient interaction process.

Thus, we hypothesized the following:

> *H1: Privacy protection settings have a positive impact on a doctor's interactive professional health care knowledge sharing in OHCs.*

Conflicts Between Privacy Protection and Searching Professional Health Care Knowledge Sharing

In addition to interactive professional health care knowledge sharing, searching professional health care knowledge sharing is another approach by which doctors can share knowledge in OHCs. As there are numerous visible consultation records in OHCs that are available to other users, it is easy for a patient to access these consultation records via search engines (eg, Bing, Google, or Yahoo!) [29]. By learning from the professional health care knowledge that doctors share with other patients in their records, patients can internalize such knowledge as their professional health care knowledge.

For doctors, searching professional health care knowledge sharing is a more widely used method of disseminating professional knowledge than interactive professional health care knowledge sharing. Compared with interactive professional health care knowledge sharing in one-to-one communication, searching professional health care knowledge can be extended to any patient who is willing to share such knowledge through the internet under the authorization of doctors. Thus, searching professional health care knowledge sharing provides a more convenient knowledge sharing channel for doctors. According to statistics, among the patients who use the internet to obtain professional health care knowledge, 61% use search engines to access searching professional health care knowledge sharing, which attracts twice the number of users who use social platforms and online communities [3].

Searching professional health care knowledge sharing, which is mainly recorded in doctor-patient interactive history records, often contains patients' specific symptoms, diagnosis results, examination records, and other matters protected by personal privacy. Although these history records will help other patients acquire searching professional health care knowledge sharing more accurately, if such private information is used indiscreetly, it will have adverse effects on those whose privacy has been invaded [16]. Therefore, the OHC provides a function for doctors to protect the privacy of their patients by making the consultation process invisible. It helps to eliminate the privacy exposure of personal health care information and thus fundamentally eliminates privacy invasion. However, while privacy is protected, other patients are not able to access the complete interaction records and, thus, are not able to clearly grasp the details of the shared professional health care knowledge. This negatively affects the searching professional health care knowledge sharing of other patients.

Thus, we hypothesized the following:

> *H2: Privacy protection settings have a negative impact on a doctor's searching professional health care knowledge sharing in OHCs.*

The Moderating Effects of Disease Stigma

Closely related to privacy protection in OHCs is the inherent stigma ascribed to many health conditions [30]. Social stigma refers to negative feelings toward an individual or a group on socially characteristic grounds that distinguish the individual or group from others [31,32].

Compared to offline treatments, online professional health care knowledge sharing provides natural protection, like a veil, for patients with diseases they may feel embarrassed about. Through an OHC, patients with stigma disorders can access professional health care knowledge sharing with doctors without the need for face-to-face interactions. In particular, people living with stigmatized conditions such as mental illness, cancer, or HIV are more likely to seek support and health information online due to the perceived anonymity the internet provides [33-35]. In their study, Zhang et al [16] demonstrated that the internet could support the disclosure of stigmatized illnesses such as HIV by helping the afflicted overcome the aspects of social stigma.

Different types of diseases possess different degrees of stigma. For instance, patients with HIV will be more worried about and feel more ashamed of other people's discrimination than patients with a cold. Berger et al [35] showed that compared to those with nonstigmatized diseases, those with stigmatized illnesses were more likely to be sensitive to privacy disclosure. This is because once a stigmatized patient's privacy is exposed, the disclosure will cause the patient more considerable pain and annoyance than for a nonstigmatized patient. Therefore, when patients with stigmatized diseases use OHCs to interact with doctors to search for professional health care knowledge sharing, they expect their privacy to be fully protected. Patients with stigmatized diseases, who are more sensitive about protecting their privacy, are more likely to trust doctors who devote attention to protecting their patients' privacy. As a consequence, their frequency of searching professional health care knowledge sharing increases.

Thus, we hypothesized the following:

> *H3: The stigma associated with a disease positively moderates the effects of privacy protection settings on interactive and searching professional health care knowledge sharing.*

Methods

Research Settings

One of the biggest OHCs in China, Good Doctor (http://www.haodf.com/), which was launched in September 2006, was chosen for our research context. It provides a virtual online platform for patients and offline registered doctors. Through Good Doctor, patients can engage in professional health care knowledge sharing with doctors through online consultations, telephone consultations, reading articles published by doctors, browsing other patient-doctor consultation records, and so on. These professional health care knowledge sharing processes may be recorded by Good Doctor, and some of them are shown to the public on doctors' profile pages. The professional health care knowledge sharing interaction records include telephone and online consultation records (ie, clinical and academic titles and detailed content of the consultation, from the doctor and patient, during the whole consultation process) with anonymous patients, the number and content of articles published by the doctor (ie, academic papers, summary of medical experience, and related knowledge forwarding), and the total number of patients who have engaged in professional health care knowledge sharing with the doctor. However, if the doctor has set privacy protection on an online consultation, the consultation details are not rendered publicly available. We developed software that crawled doctors' profile pages in Good Doctor every month from January 2008 to April 2016. Finally, we constructed a monthly panel data set for 19,456 doctors.

Variables

Professional health care knowledge sharing was measured by the number of patients who interacted with a doctor or searched the consultation records of that doctor. The number of online consultations between patients and a doctor through which they engage in professional health care knowledge sharing interaction was counted based on the doctor's consultation list and was used to measure our first dependent variable: the doctor's interactive professional health care knowledge sharing (intrctPHKS). The second dependent variable, searching professional health care knowledge sharing (searchPHKS), indicates the number of patients who learned of a doctor's professional health care knowledge through searching that doctor's consultation records without interacting with the doctor. Although this number cannot be collected directly from the website, it can be inferred from the calculation of the number of patients who visit the doctor's profile page minus the number of patients who interact via consultations. Even though the number of patients visiting the profile page may count in "noise" values (such as browsing a doctor's personal information, or repeat visits), in the difference-in-difference model, a random error in the dependent variable will not influence regression estimates of treatment effect. Hence, by ignoring the cases of a patient viewing a doctor's profile page unintentionally, it is assumed that all patients visiting a doctor's page have access to professional health care knowledge sharing with this doctor. In other words, in addition to the number of patients interacting with a doctor through consultations, patients who visit a doctor's page have access to searching professional health care knowledge sharing with that doctor.

Online consultations between doctors and patients can be conducted by means of text and images, and the records of consultations are open to the public unless the doctor changes the privacy protection settings of the records. During a doctor-patient consultation, the doctor has the authority to choose whether to make the consultation records of their patient visible to the public. If an online consultation is set as private, only the patient and the attending doctor can access the entire interaction history. Others who visit this doctor's profile page can view only the topic of a private consultation without seeing the detailed information. We calculated both the total number of a doctor's online consultations with privacy protection settings and the monthly number of privacy protection settings to measure our independent variable: PrivacySetting.

Moreover, patients with different levels of disease stigma may have different attitudes toward privacy sensitivity regarding online professional health care knowledge sharing interactions with doctors. To understand how the stigma of diseases affects the privacy protection setting of consultations on interactive and searching professional health care knowledge sharing, disease stigma (Stigma) was considered as the moderate effect of our model. We used the classification method of De Choudhury et al [8] to measure stigma on a 2-point scale: 1=high stigma and 0=low stigma.

To account for potential confounding effects that influence professional health care knowledge sharing, we included other control variables (ie, the doctor's title, virtual gifts, articles posted online, and thank-you letters). According to the literature, a doctor's offline status and online reputation will also affect their knowledge sharing in OHCs [36]. Therefore, the variables that represent the doctor's offline status and online reputation are treated as our control variables (ie, the title of a doctor [Title], the number of virtual gifts [Gift], the number of thank-you letters [Letter], and the number of published articles [Article]). The official clinical title certified by China's national agency based on uniform standards was used to measure Title. In general, 4 rankings exist for titles: Fellow (4), Associate Fellow (3), Attending (2), Resident (1), and none (0). The variable Gift was used to represent the number of virtual gifts sent by the patients who consulted online. Thank-you letters (Letter) denote the number of thank-you letters submitted by patients who had visited a doctor in both the online and offline contexts. The number of articles the doctor had published online was counted as Article. The defined variables are shown in Table 1.

Descriptive Statistics

Our monthly panel data set included the doctor's identification number, the number of online consultations with and without privacy protection settings, the total number of visits, the number of patients, the number of articles, the number of gifts, the number of follow-up visits, and the title of the doctor. We obtained 631,529 online consultations of 19,456 doctors. More detailed descriptive statistics are shown in Table 2.

The correlation of each indicator using a cross-section is shown in Table 3, which indicates values in the acceptable range. The correlation coefficients between the independent variables are relatively small. In other words, the possibility of variable redundancy is relatively small.

Table 1. Variables defined.

Variable	Operational definition
Dependent	
intrctPHKS[a]	The number of interactive professional health care knowledge sharing
searchPHKS[b]	The number of searching professional health care knowledge sharing
Independent	
PrivacySetting	The number of online consultations with a privacy protection setting
Moderate	
Stigma	The stigma of a disease for which a doctor offers consultations
Control	
Gift	The number of virtual gifts received by a doctor
Letter	The number of thank-you letters
Title	The title of the doctor
Article	The number of articles written by the doctor

[a]intrctPHKS: interactive professional health care knowledge sharing.

[b]searchPHKS: searching professional health care knowledge sharing.

Table 2. Descriptive statistics for selected variables.

Variable	Observed	Mean (SD)	Minimum	Maximum
Dependent				
intrctPHKS[a]	19,456	459.9 (1528)	0	64,561
searchPHKS[b]	19,456	333,510 (1,420,069)	7	62,300,000
Independent				
PrivacySetting	19,456	0.133 (1.106)	0	23
Moderate				
Stigma	19,456	0.403 (0.491)	0	1
Control				
Gift	19,456	20.67 (96.87)	0	2996
Letter	19,456	6.723 (19.26)	0	382
Title	19,456	3.895 (0.923)	1	5
Article	19,456	8.920 (46.82)	0	2158

[a]intrctPHKS: interactive professional health care knowledge sharing.

[b]searchPHKS: searching professional health care knowledge sharing.

Table 3. Correlation coefficients between independent variables.

Variable	intrctPHKS[a]	searchPHKS[b]	PrivacySetting	Stigma	Gift	Letter	Title	Article
intrctPHKS	1.000	__[c]	—	—	—	—	—	—
searchPHKS	0.904	1.000	—	—	—	—	—	—
PrivacySetting	0.097	0.099	1.000	—	—	—	—	—
Stigma	0.016	0.006	0.014	1.000	—	—	—	—
Gift	0.733	0.688	0.081	0.034	1.000	—	—	—
Letter	0.676	0.606	0.086	0.050	0.704	1.000	—	—
Title	0.141	0.130	0.011	0.110	0.115	0.202	1.000	—
Article	0.302	0.366	0.051	0.039	0.211	0.184	0.074	1.000

[a]intrctPHKS: interactive professional health care knowledge sharing.

[b]searchPHKS: searching professional health care knowledge sharing.

[c]N/A: not applicable.

Empirical Models

Observing the privacy protection setting of doctors over various periods creates a natural experimental setting that allows a comparison of the effects of differences in privacy protection settings before and after a doctor uses this function. Constructing a panel data set of doctors' monthly online consultations, we estimated the difference-in-difference models, reflected by equations (1) through (4). Difference-in-difference models were used to calculate the effect of a treatment on an outcome by comparing the outcome average change over time from the control group to the treatment group. It helps to remove biases of permanent differences between the treatment and control groups, as well as trend biases caused by the changing of other factors of the outcome over time [37]. Thus, we exploited the effect of privacy protection settings on professional health care knowledge sharing by identifying an exogenous variation of professional health care knowledge sharing that the privacy protection settings are set by different doctors in different months. This identification strategy has been implemented in several extant studies (eg, in Chan and Ghose [38]). Our estimation incorporated doctor-level fixed effects, which allowed us to effectively control for doctor-level unobserved heterogeneity. To control the trend effect of each doctor's professional health care knowledge sharing, we added the time fixed effect in all 4 models. Moreover, we added the individual fixed effect to control the omitted variable bias caused by unobserved heterogeneity. Both the time fixed effect and the individual fixed effect were set as dummy variables. Furthermore, we used a cross-sectional data set to conclude our analyses with a set of robustness checks. Accordingly, our empirical models of doctors' privacy protection settings in OHCs on both their interactive and searching professional health care knowledge sharing are as follows:

$$\text{intrctPHKS}_{it} = \beta_0 + \beta_1 \text{PrivacySetting}_{it} + \beta_2 \text{Stigma}_i + \beta_3 C + A_i + T_t + \varepsilon_{it} \quad (1)$$

$$searchPHKS_{it} = \beta_0 + \beta_1 PrivacySetting_{it} + \beta_2 Stigma_i + \beta_3 C + A_i + T_t + \varepsilon_{it} \quad (2)$$

The moderate effect models are as follows:

$$intrctPHKS_{it} = \beta_0 + \beta_1 PrivacySetting_{it} + \beta_2 Stigma_i + \beta_7 Stigma_{it}*PrivacySetting_{it} + \beta_3 C + A_i + T_t + \varepsilon_{it} \quad (3)$$

$$searchPHKS_{it} = \beta_0 + \beta_1 PrivacySetting_{it} + \beta_2 Stigma_i + \beta_7 Stigma_{it}*PrivacySetting_{it} + \beta_3 C + A_i + T_t + \varepsilon_{it} \quad (4)$$

Where *i* indexes indicate the doctor and *t* indexes indicate time (monthly). A_i is a vector of doctor fixed effects; T_t is a vector of time fixed effects; $PrivacySetting_{it}$ is the binary indicator for the privacy protection setting (that is, $PrivacySetting_{it}$=1 if the doctor has set the privacy in a particular month; zero otherwise); and ε_{it} is an error term. The coefficient β_1 is the difference-in-difference estimate of the effect of a privacy protection setting. If β_1>0, then the privacy protection setting has caused an increase in professional health care knowledge sharing. Control variables are indicated by *C*.

Results

Hypothesis Testing

Stata (StataCorp) was used for our statistical analysis, and the results are shown in Table 4. The adjusted R^2 (0.037, –0.195, 0.037, and –0.223 for models 1, 2, 3, and 4 respectively) and *F* values ($F_{101,143415}$=384.8419, $F_{102,143414}$=381.0676, $F_{101,143376}$=34.43728, and $F_{3,143474}$=5.534647 for models 1, 2, 3, and 4 respectively) were reasonable and significant. The results of the variance inflation factor statistics for the variable indicated no multicollinearity (the variance inflation factor statistic of every variable is not greater than 2.0). The results of models only with our main treatment effect (PrivacySetting) are shown in columns 1 and 3 for interactive professional health care knowledge sharing and searching professional health care knowledge sharing, respectively. The results of models added to the moderate effects of stigma are shown in columns 2 and 4 for interactive professional health care knowledge sharing and searching professional health care knowledge sharing, respectively.

Hypothesis 1 postulated that the privacy protection setting has a significant positive effect on interactive professional health care knowledge sharing. In column 1 of Table 4, this hypothesis is supported as the coefficient of privacy protection setting was seen as positive and statistically significant (β_1=.123, *P*<.001). However, we observed in column 3 a significantly negative effect of the privacy protection setting on doctors' searching professional health care knowledge sharing (β_1=–.225, *P*=.050). Therefore, hypothesis 2, which posits that the privacy protection setting has a negative effect on searching professional health care knowledge sharing online, is also supported. As both the dependent variables of model 1 (intrctPHKS) and model 3 (searchPHKS) represent the amount of a doctor's knowledge sharing, a Wald test [39] was applied to the coefficients for further comparison. In the results for PrivacySetting, the absolute value of the coefficient of searching professional health care knowledge sharing was significantly larger than that of interactive professional health care knowledge sharing (Wald test, $F_{2,175548}$=324.21; *P*<.001).

Hypothesis 3 investigated the moderate impact of disease stigma on the relation between privacy protection settings and professional health care knowledge sharing. Columns 2 and 4 of Table 4 show the stable significant result of the positive moderate effect of disease stigma. In other words, high disease stigma positively impacts the effects of privacy protection settings on interactive professional health care knowledge sharing (coefficients are in the same valence) and negatively impacts the effects of privacy protection settings on searching professional health care knowledge sharing (coefficients are in the reverse valence). Therefore, hypothesis 3 is verified. Specifically, for disease stigma, we found that the stigma has a positive correlation with professional health care knowledge sharing.

Robust Check

To check the robustness of our results, we conducted a cross-sectional rerun of the model. The results presented in Table 5 are consistent with the results of the previous model. The results show that the treatment and the moderate effect have a significant positive effect, which is the same as the main results.

All the control variables have a statistically significant effect on professional health care knowledge sharing. Doctors with higher online and offline reputations will attract more patients [40]. The title reflects a doctor's offline reputation, and higher-level doctors are known to be more professional. Hence, the title has a positive effect on doctors' professional health care knowledge sharing. Moreover, virtual gifts, the number of articles, and thank-you letters can be proxy variables of a doctor's online reputation, which can positively affect professional health care knowledge sharing.

Table 4. Econometric models of the treatment effect and moderate effect. All the control variables and the individual and time fixed effect variables were included in the 4 models. The interaction terms (PrivacySetting*Stigma) represent the moderate effect of stigma.

Variable	intrctPHKS[a] (N=175,548)				searchPHKS[b] (N=175,509)			
	Model 1		Model 2		Model 3		Model 4	
	Treatment effect		Moderate effect		Treatment effect		Moderate effect	
	β (SE)[c]	*P* value	β (SE)	*P* value	β (SE)	*P* value	β (SE)	*P* value
PrivacySetting	.123 (.063)	<.001	.105 (.084)	.009	–.225 (.024)	.05	–.224 (.032)	<.001
Stigma	.024 (.005)	<.001	.024 (.005)	<.001	.033 (.005)	.008	.044 (.005)	.003
PrivacySetting*Stigma	__[d]	—	.041 (.125)	.08	—	—	.089 (.049)	.07
Constant	2.665 (3.482)	.44	2.665 (3.482)	.44	.414 (1.335)	.76	.376 (.003)	.001
Control	YES	N/A[e]	YES	N/A	YES	N/A	YES	N/A
Individual fixed effect	YES	N/A	YES	N/A	YES	N/A	YES	N/A
Time fixed effect	YES	N/A	YES	N/A	YES	N/A	YES	N/A
R^2	.037	N/A	.037	N/A	–0.195	N/A	–.223	N/A

[a]intrctPHKS: interactive professional health care knowledge sharing.

[b]searchPHKS: searching professional health care knowledge sharing.

[c]Values in parentheses are the robust standard errors.

[d]Not available. Models 1 and 3 did not contain the moderate effect of stigma.

[e]N/A: not applicable.

Table 5. Results of the robust check using cross-section data (N=23,112 for all models).

Variable	intrctPHKS[a]				searchPHKS[b]			
	Model 1		Model 2		Model 3		Model 4	
	β (SE)[c]	*P* value	β (SE)	*P* value	β (SE)	*P* value	β (SE)	*P* value
PrivacySetting	.0196 (4.35)	<.001	.0277 (4.71)	<.001	–.030 (.012)	.05	–.032 (.013)	.02
Stigma	.0308 (6.86)	<.001	.0297 (6.55)	<.001	.032 (0.005)	.002	.032 (0.005)	.004
PrivacySetting*Stigma	__[d]	—	.0126 (2.14)	.06	—	—	.018 (.009)	.02
Title	.0158 (3.44)	.001	.0160 (3.48)	.006	.017 (.006)	<.001	.017 (.006)	<.001
Article	.148 (32.20)	<.001	.148 (32.20)	<.001	.222 (.058)	<.001	.222 (.058)	.001
Gift	.490 (77.11)	<.001	.490 (77.04)	<.001	.220 (.033)	<.001	.220 (.033)	<.001
Letter	.297 (46.16)	<.001	.297 (46.17)	<.001	.483 (.055)	<.001	.483 (.055)	<.001
Constant	1.91e-10 (.00)	>.99	1.33e-10 (.00)	.54	–.000 (.004)	.32	.000 (.004)	.83
R^2	.6081	N/A[e]	.6081	N/A	.552	N/A	.552	N/A

[a]intrctPHKS: interactive professional health care knowledge sharing.

[b]searchPHKS: searching professional health care knowledge sharing.

[c]Values in parentheses are the robust standard errors.

[d]Not available. Models 1 and 3 did not present the moderate effect of stigma.

[e]N/A: not applicable.

Discussion

Principal Findings

This study examined the role of a doctor's privacy protection settings in OHCs on professional knowledge sharing. According to the data set from a distinguished OHC in China, both interactive professional health care knowledge sharing (intrctPHKS) and searching professional health care knowledge sharing (searchPHKS) are considered to be representative of a doctor's professional health care knowledge sharing in OHCs. The stigma of diseases is estimated to moderate the association of privacy protection settings and professional health care knowledge sharing.

Our results indicate that our hypotheses on the effect of doctors' privacy protection settings on professional health care

knowledge sharing are significantly supported. The more privacy protection setting a doctor establishes, the more interactive their professional health care knowledge sharing will be. This means that if a doctor sets their interaction records as private, their patients will prefer to make multiple interactions, and more patients will consult this doctor. Thus, a higher performance of interactive professional health care knowledge sharing can be achieved. Similar studies in other privacy-sensitive contexts, such as in financial loans, e-commerce, and so on, have demonstrated that better privacy protection mechanisms increase user trust and, thus, increase user willingness to disclose personal information for better management [41,42].

However, a doctor's privacy protection setting will limit the number of patients searching for professional health care knowledge online. This may lead to patients needing more time to find the right doctor, or they might not learn some useful information on the same health conditions. Therefore, doctors' privacy protection settings restrict their searching for professional health care knowledge sharing with patients. Switching off privacy protection settings properly is good for increasing searching professional health care knowledge sharing. This result is also valid in financial fields, where the appropriate sharing of privacy can overcome the information asymmetry of financial credit [43]. However, it is more acceptable for OHC users to disclosure their online consultation records for searching professional health care knowledge sharing than for borrowers to disclose their financial records for overcoming the information asymmetry. This may be due to the nature of professional health care knowledge sharing, which is reciprocal, and disclosure behaviors can give patients a sense of achievement. However, financial privacy is related to personal property, which a person is usually unwilling to share [44]. The difference has also been confirmed by Xu et al [45] that privacy sensitivity varies across different types of websites (eg, in e-commerce, financial, and health care).

As privacy protection settings positively affect interactive professional health care knowledge sharing and negatively affect searching professional health care knowledge sharing, it presents a bilateral switch of professional health care knowledge sharing, which is positive on one side and must at the same time be negative on the other side. The doctor who prefers to establish privacy protection will enjoy better performance in interactive professional health care knowledge sharing, but this weakens the performance of professional health care knowledge sharing searching. Conversely, a doctor who prefers public records will have better searching professional health care knowledge sharing and relatively weak interactive professional health care knowledge sharing. Figure 1 illustrates the metaphor of the bilateral switch. A doctor's professional health care knowledge sharing is a process of professional health care knowledge flowing from doctors to patients, like blood flowing from one side to the other side in blood vessels. In Figure 1, the privacy protection setting is likened to a bilateral switch that is installed in the blood vessels to regulate blood shunting. When blood (professional health care knowledge sharing) meets the switch (privacy protection settings), a bifurcation moves upward (interactive professional health care knowledge sharing), and the other bifurcation flows downward (searching professional health care knowledge sharing). If a doctor creates a stronger privacy protection setting, the bilateral switch of professional health care knowledge sharing is a downward regulation, thus resulting in more of interactive professional health care knowledge sharing and less of searching professional health care knowledge sharing. However, when the doctor prefers weaker privacy protection settings, there will be more searching professional health care knowledge sharing and less interactive professional health care knowledge sharing, with the switch turning up.

However, as we noted previously, according to the Wald test, the absolute value of β_1 in searching professional health care knowledge sharing is higher than that in interactive professional health care knowledge sharing. Therefore, when the same number of consultations are added that have a privacy protection setting, the loss of searching professional health care knowledge sharing is greater than that of the interactive professional health care knowledge sharing. In other words, the bilateral switch of professional health care knowledge sharing does not form a straight line as shown in Figure 1, but instead shows a thick switch that may cause a loss of professional health care knowledge sharing. We show an example in Figure 2.

In addition, empirical results indicate that the moderating effects of stigma positively affect the relationship between privacy protection settings and interactive professional health care knowledge sharing, and have a negative influence on the relationship between privacy protection settings and searching professional health care knowledge sharing. In other words, for high-stigma diseases, the increase in privacy protection settings is better for interactive professional health care knowledge sharing, while at the same time weakening the reduction of the searching professional health care knowledge sharing. This can be visualized as the stigma adjusting the thickness of the switch, namely professional health care knowledge sharing losses. The higher the stigma, the smaller the thickness of the switch (less loss). Conversely, the lower the stigma, the higher the thickness of the switch (more loss). This is because the virtual interaction mechanism of OHCs appears to be more acceptable by patients with high disease stigma. We show an example in Figure 3. OHCs recognize that they would have more visitors if they provide professional disease treatment support, as patients may disregard the disease stigma.

Figure 1. Illustration of the effect of privacy protection settings on interactive and searching professional health care knowledge sharing. PHKS: professional health care knowledge sharing.

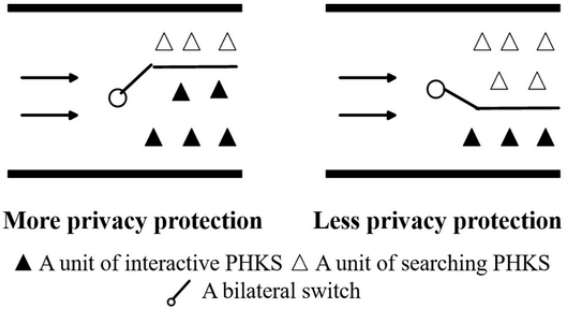


Figure 2. Illustration of the effect of privacy protection settings on interactive and searching professional health care knowledge sharing in an ideal and actual situation. PHKS: professional health care knowledge sharing.

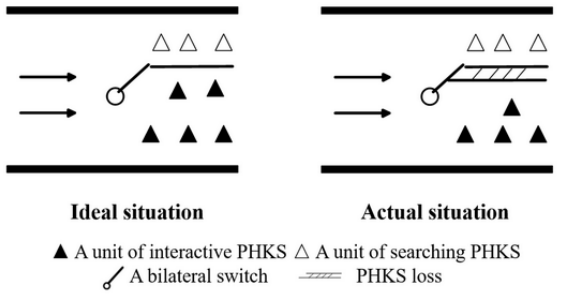
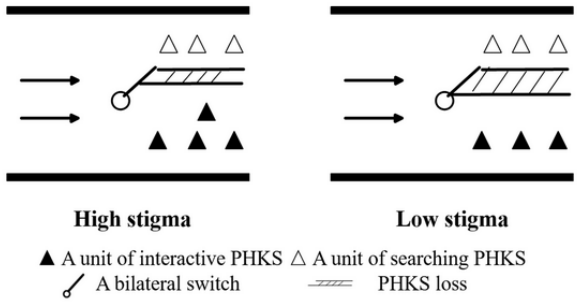


Figure 3. Illustration of the effect of disease stigma on interactive and searching professional health care knowledge sharing. PHKS: professional health care knowledge sharing.



Theoretical Implications

This study augments OHC-related research with several insights. First, the OHC, as a new application of a sharing economy, provides an efficient sharing platform for scarce professional health care knowledge [46]. Like in other forms of sharing economies, the opportunity for OHCs to share valuable professional health care knowledge is enormous. However, compared with other sharing economy applications, the OHC also faces privacy protection challenges due to the sensitivity and life-threatening nature of medical information. Although previous studies have studied privacy-protection technologies, patient privacy concerns, and intentions, the current OHC privacy protection mechanism is still incomplete. The privacy protection mechanism established by doctors was in an experimental stage in the target website of this study (Good Doctor). Privacy protection settings by doctors have the advantage that doctors can control whether to install privacy protection settings from a professional and objective perspective (compared with patients' subjective perspectives). Therefore, we proposed a research framework that verifies the effect of a doctor's privacy protection mechanism on professional health care knowledge sharing.

Second, we abstracted knowledge sharing in an OHC into interactive professional health care knowledge sharing and searching professional health care knowledge sharing, and empirically proved the significant differences in privacy protection settings between them. A privacy protection setting, like a bilateral switch of professional health care knowledge sharing, positively affects interactive professional health care knowledge sharing and negatively affects searching professional health care knowledge sharing. In addition to the positive effects of privacy protection settings, we uncovered the paradox between privacy protection settings and knowledge diffusion. In other words, the effects of privacy protection settings on knowledge sharing were not seen to be stable and required a balanced state of a bilateral switch. This is another paradox of privacy protection setting in addition to the privacy-personalization paradox [47], which can be denominated as the privacy-knowledge-sharing paradox.

We also examined the moderating effects of disease stigma on the association between privacy protection settings and doctors' online professional health care knowledge sharing. Social stigma has been studied widely in the privacy field but rarely examined in existing studies on knowledge sharing. The empirical studies considered the direct or indirect influences of stigma on an individual's intention to disclose information [21]. However, there is a lack of empirical studies that examine the direct or moderating effect of disease stigma on knowledge sharing. We determined that disease stigma has a positive moderating effect on the relationship of privacy protection settings with online professional health care knowledge sharing. In particular, this result indicates a facilitating effect of privacy protection settings on interactive professional health care knowledge sharing and an inhibiting effect of privacy protection settings on searching professional health care knowledge sharing.

Practical Implications

This study provides several important practical implications for both OHC administrators and doctors. Our results indicate that a doctor's privacy protection settings are the bilateral switch of professional health care knowledge sharing. For OHC managers, protecting patient privacy helps to improve the participation and continuous use intentions of patients. However, it also hinders the extensive spreading of professional health care knowledge, which can help users to realize the social value of the OHC and increase its influence. For doctors, their privacy protection settings help to reduce patient privacy concerns and thus encourage patients to provide more accurate information to provide better advice. In addition, the privacy protection setting renders the doctor's professional health care knowledge more difficult for patients to access in a search, thus reducing the doctor's "exposure" in the OHCs. The fewer patients the doctors have, the less knowledge they are likely to share. Therefore, we postulate that privacy protection settings are a bilateral switch of professional health care knowledge that needs balancing between OHC managers and doctors. The question is how to balance such a bilateral switch, which impacts the performance of both interactive professional health care knowledge sharing and searching professional health care knowledge sharing. Moreover, our empirical results improve the decision making of OHC managers and doctors. Specifically, the privacy protection setting has a cutoff effect on searching professional health care knowledge sharing and a promoting effect on interactive professional health care knowledge sharing. Furthermore, the cutoff effect is weaker than the promoting effect. Therefore, OHC managers and doctors can make their decision (by toggling the bilateral switch of professional health care knowledge sharing) according to whether their target is the OHC or the individual propensity. In addition, the moderating effect of disease stigma positively affects the relation between privacy protection settings and professional health care knowledge sharing. In other words, for a high-stigma disease, the professional health care knowledge sharing process is more friendly toward privacy protection settings. However, a low-stigma disease needs less privacy protection. This will help managers of OHCs and doctors to balance the bilateral switch of professional health care knowledge sharing based on disease stigma.

Limitations and Future Research Directions

Our paper has some limitations that can serve as potential areas for future research. First, our research focuses on the OHC context. The methodologies and insights have the potential to be generalized to other wellness-care contexts, such as mHealth. However, examining the mHealth privacy paradox to compare the relationship and heterogeneity of the impact on patient behavior and outcomes under different health care contexts would be more beneficial and vital for future research.

Second, despite the collection of data from a typical OHC in China and the provision of a unique data set with doctors' privacy protection settings, the general applicability of our conclusions may be limited. Future research could extend this study across different OHCs and cultural settings to enhance the generalizability of our study.

Additionally, although our results effectively explain the privacy protection settings of doctors on professional health care knowledge sharing in OHCs, in-depth interviews or surveys with patients could further investigate perceptions of patients on the relationship between privacy protection settings and professional health care knowledge sharing. Moreover, given the important role of patients in the online professional health care knowledge sharing process, in the future, researchers could also explore how patient-related variables (ie, health status, health conditions, disease severity, participant motivations and reasons) affect doctors' online professional health care knowledge sharing with patients. Although both the difference-in-difference models (in our main analysis) and the cross-sectional models (in the robust check) showed significant verification of our hypotheses, they also indicate that the control variables may have strong explanatory power on doctors' professional health care knowledge sharing (ie, gift 0.49, $P<.001$). These control variables may be the proxy variables for the quality of doctors' treatment. Therefore, future research can deeply explore the effect of online response quality on professional health care knowledge sharing.

Conclusion

The important role of OHCs in sharing health care knowledge has been widely acknowledged by both practitioners and researchers; however, significant gaps remain in our exploration of the privacy protection mechanism in OHCs. This study examined how doctors' privacy protection setting choices affect their professional health care knowledge sharing in a Chinese OHC, a topic that, for the most part, has not been studied from the perspective of knowledge sharing. We used a fresh category of professional health care knowledge sharing; namely, interactive professional health care knowledge sharing and searching professional health care knowledge sharing. Doctors play different roles in these two kinds of professional health care knowledge sharing, both as active and passive actors, where different knowledge sharing processes occur. Using an empirical study on a large OHC, this study identified the bilateral role of privacy protection settings on professional health care knowledge sharing. It is a privacy-knowledge–sharing paradox in which the professional health care knowledge sharing process is not stable and calls for a balanced state. Our findings offer implications for doctors to gain a better understanding of how

to set privacy protection for better professional health care knowledge sharing. Different privacy protection mechanisms can be suggested by OHC executives, which are customized to suit each level of disease stigma. We hope that this study promotes additional research to further enrich our understanding of professional health care knowledge sharing in OHCs.

Acknowledgments

This study was partially funded by the National Natural Science Foundation of China (71801062, 71871073, 71871074, 71531007), China Postdoctoral Science Foundation (2018M640301, 2019T120278), and the Key Projects of Philosophy and Social Sciences Research of Chinese Ministry of Education (19JZD021).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Further previous research.
[DOC File , 98 KB - jmir_v22i5e16246_app1.doc]

References

1. Sood S, Mbarika V, Jugoo S, Dookhy R, Doarn CR, Prakash N, et al. What is telemedicine? A collection of 104 peer-reviewed perspectives and theoretical underpinnings. Telemed J E Health 2007 Oct;13(5):573-590. [doi: 10.1089/tmj.2006.0073] [Medline: 17999619]
2. Eysenbach G, Köhler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. BMJ 2002 Mar 09;324(7337):573-577 [FREE Full text] [doi: 10.1136/bmj.324.7337.573] [Medline: 11884321]
3. Fox S. Pew Research Center. The social life of health information 2011: Pew Internet & American Life Project Washington, DC; 2011. The Social Life of Health Information, 2011 URL: https://www.pewresearch.org/internet/2011/05/12/the-social-life-of-health-information-2011/ [accessed 2020-04-22]
4. Ayers SL, Kronenfeld JJ. Chronic illness and health-seeking information on the Internet. Health (London) 2007 Jul 24;11(3):327-347. [doi: 10.1177/1363459307077547] [Medline: 17606698]
5. Hawn C. Take two aspirin and tweet me in the morning: how Twitter, Facebook, and other social media are reshaping health care. Health Aff (Millwood) 2009 Mar;28(2):361-368. [doi: 10.1377/hlthaff.28.2.361] [Medline: 19275991]
6. Gruber TR. Toward principles for the design of ontologies used for knowledge sharing? International Journal of Human-Computer Studies 1995 Nov;43(5-6):907-928. [doi: 10.1006/ijhc.1995.1081]
7. Qun Z, Xiaocheng Z. The Design of Individual Knowledge Sharing Platform Based on Blog for Online Information Literacy Education. Physics Procedia 2012;33:1426-1432. [doi: 10.1016/j.phpro.2012.05.233]
8. De Choudhury M, Ringel Morris M, White R. Seeking and sharing health information online: comparing search engines and social media. 2014 Presented at: Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems; April 2014; Toronto p. 1365-1376. [doi: 10.1145/2556288.2557214]
9. Johnson GJ, Ambrose PJ. Neo-tribes. Commun. ACM 2006 Jan 01;49(1):107-113. [doi: 10.1145/1107458.1107463]
10. Skeels M, Unruh K, Powell C, Pratt W. Catalyzing social support for breast cancer patients. 2010 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; April 2010; Atlanta p. 173-182. [doi: 10.1145/1753326.1753353]
11. Eysenbach G, Powell J, Englesakis M, Rizo C, Stern A. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. BMJ 2004 May 15;328(7449):1166 [FREE Full text] [doi: 10.1136/bmj.328.7449.1166] [Medline: 15142921]
12. Oh HJ, Lauckner C, Boehmer J, Fewins-Bliss R, Li K. Facebooking for health: An examination into the solicitation and effects of health-related social support on social networking sites. Computers in Human Behavior 2013 Sep;29(5):2072-2080. [doi: 10.1016/j.chb.2013.04.017]
13. Cartright MA, White R, Horvitz E. Intentions and attention in exploratory health search. 2011 Presented at: Proceedings of the 34th international ACM SIGIR Conference on Research and Development in Information Retrieval; 2011; Beijing p. 65-74. [doi: 10.1145/2009916.2009929]
14. Yan L, Tan Y. Feeling Blue? Go Online: An Empirical Study of Social Support Among Patients. Information Systems Research 2014 Dec;25(4):690-709. [doi: 10.1287/isre.2014.0538]
15. Fox S, Duggan M. Pew Research Center. 2013 Jan 13. Health Online 2013 URL: https://www.pewresearch.org/internet/2013/01/15/health-online-2013/ [accessed 2020-04-22]

16. Zhang X, Liu S, Chen X, Wang L, Gao B, Zhu Q. Health information privacy concerns, antecedents, and information disclosure intention in online health communities. Information & Management 2018 Jun;55(4):482-493. [doi: 10.1016/j.im.2017.11.003]
17. Li J. Improving chronic disease self-management through social networks. Popul Health Manag 2013 Oct;16(5):285-287. [doi: 10.1089/pop.2012.0110] [Medline: 23672233]
18. Moubarak G, Guiot A, Benhamou Y, Benhamou A, Hariri S. Facebook activity of residents and fellows and its impact on the doctor-patient relationship. J Med Ethics 2011 Feb 15;37(2):101-104. [doi: 10.1136/jme.2010.036293] [Medline: 21160080]
19. Thompson LA, Black E, Duff WP, Paradise Black N, Saliba H, Dawson K. Protected health information on social networking sites: ethical and legal considerations. J Med Internet Res 2011 Jan 19;13(1):e8 [FREE Full text] [doi: 10.2196/jmir.1590] [Medline: 21247862]
20. Li J. Privacy policies for health social networking sites. J Am Med Inform Assoc 2013 Jul 01;20(4):704-707 [FREE Full text] [doi: 10.1136/amiajnl-2012-001500] [Medline: 23599228]
21. Bansal G, Zahedi F, Gefen D. The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. Decision Support Systems 2010 May;49(2):138-150. [doi: 10.1016/j.dss.2010.01.010]
22. Farič N, Potts HWW. Motivations for contributing to health-related articles on Wikipedia: an interview study. J Med Internet Res 2014 Dec 03;16(12):e260 [FREE Full text] [doi: 10.2196/jmir.3569] [Medline: 25498308]
23. Murad A, Hyde N, Chang S, Lederman R, Bosua R, Pirotta M, et al. Quantifying Use of a Health Virtual Community of Practice for General Practitioners' Continuing Professional Development: A Novel Methodology and Pilot Evaluation. J Med Internet Res 2019 Nov 27;21(11):e14545 [FREE Full text] [doi: 10.2196/14545] [Medline: 31774401]
24. Rolls K, Hansen M, Jackson D, Elliott D. How Health Care Professionals Use Social Media to Create Virtual Communities: An Integrative Review. J Med Internet Res 2016 Jun 16;18(6):e166 [FREE Full text] [doi: 10.2196/jmir.5312] [Medline: 27328967]
25. Jing D, Jin Y, Liu J. The Impact of Monetary Incentives on Physician Prosocial Behavior in Online Medical Consulting Platforms: Evidence From China. J Med Internet Res 2019 Jul 26;21(7):e14685 [FREE Full text] [doi: 10.2196/14685] [Medline: 31350834]
26. Westin AF. Privacy and Freedom. Michigan Law Review 1968 Mar;66(5):1064-1075. [doi: 10.2307/1287193]
27. Esmaeilzadeh P. The Impacts of the Perceived Transparency of Privacy Policies and Trust in Providers for Building Trust in Health Information Exchange: Empirical Study. JMIR Med Inform 2019 Nov 26;7(4):e14050 [FREE Full text] [doi: 10.2196/14050] [Medline: 31769757]
28. Li J. A Privacy Preservation Model for Health-Related Social Networking Sites. J Med Internet Res 2015 Jul 08;17(7):e168 [FREE Full text] [doi: 10.2196/jmir.3973] [Medline: 26155953]
29. Rice RE. Influences, usage, and outcomes of Internet health information searching: multivariate results from the Pew surveys. Int J Med Inform 2006 Jan;75(1):8-28. [doi: 10.1016/j.ijmedinf.2005.07.032] [Medline: 16125453]
30. Corrigan P. How stigma interferes with mental health care. American Psychologist 2004 Oct;59(7):614-625. [doi: 10.1037/0003-066x.59.7.614]
31. Crocker J, Major B. Social stigma and self-esteem: The self-protective properties of stigma. Psychological Review 1989;96(4):608-630. [doi: 10.1037/0033-295x.96.4.608]
32. Major B, O'Brien LT. The social psychology of stigma. Annu Rev Psychol 2005 Feb;56(1):393-421. [doi: 10.1146/annurev.psych.56.091103.070137] [Medline: 15709941]
33. Mo PK, Coulson NS. Online support group use and psychological health for individuals living with HIV/AIDS. Patient Educ Couns 2013 Dec;93(3):426-432. [doi: 10.1016/j.pec.2013.04.004] [Medline: 23669152]
34. Maestre J, Herring S, Min A, Connelly C, Shih P. Where and How to Look for Help Matters: Analysis of Support Exchange in Online Health Communities for People Living with HIV. Information 2018 Oct 20;9(10):259. [doi: 10.3390/info9100259]
35. Berger M, Wagner TH, Baker LC. Internet use and stigmatized illness. Soc Sci Med 2005 Oct;61(8):1821-1827. [doi: 10.1016/j.socscimed.2005.03.025] [Medline: 16029778]
36. Guo S, Guo X, Fang Y, Vogel D. How Doctors Gain Social and Economic Returns in Online Health-Care Communities: A Professional Capital Perspective. Journal of Management Information Systems 2017 Aug 17;34(2):487-519. [doi: 10.1080/07421222.2017.1334480]
37. Angrist J, Pischke JS. Mostly Harmless Econometrics. Princeton: Princeton University Press; 2008.
38. Chan J, Ghose A. Internet's Dirty Secret: Assessing the Impact of Online Intermediaries on HIV Transmission. MISQ 2014 Apr 4;38(4):955-976. [doi: 10.25300/misq/2014/38.4.01]
39. Mithas S, Lucas HC. Are Foreign IT Workers Cheaper? U.S. Visa Policies and Compensation of Information Technology Professionals. Management Science 2010 May;56(5):745-765. [doi: 10.1287/mnsc.1100.1149]
40. Wu H, Lu N. Online written consultation, telephone consultation and offline appointment: An examination of the channel effect in online health communities. Int J Med Inform 2017 Nov;107:107-119. [doi: 10.1016/j.ijmedinf.2017.08.009] [Medline: 29029686]
41. Jappelli T, Pagano M. Information sharing, lending and defaults: Cross-country evidence. Journal of Banking & Finance 2002 Oct;26(10):2017-2045. [doi: 10.1016/s0378-4266(01)00185-6]

42. Smith HJ, Dinev T, Xu H. Information Privacy Research: An Interdisciplinary Review. MIS Quarterly 2011;35(4):989-1015. [doi: 10.2307/41409970]
43. Pagano M, Jappelli T. Information Sharing in Credit Markets. The Journal of Finance 1993 Dec;48(5):1693-1718. [doi: 10.2307/2329064]
44. Lessig L. Privacy as property. Social Research: An International Quarterly 2002;69(1):247-269.
45. Xu H, Teo HH, Tan B. Information Privacy in the Digital Era: An Exploratory Research Framework. 2006 Presented at: AMCIS 2006 Proceedings; December 2006; Acapulco p. 899-906.
46. Liu J, Bian Y, Ye Q, Jing D. Free for Caring? The Effect of Offering Free Online Medical-Consulting Services on Physician Performance in e-Health Care. Telemed J E Health 2019 Oct 01;25(10):979-986. [doi: 10.1089/tmj.2018.0216] [Medline: 30566383]
47. Lee DJ, Ahn JH, Bang Y. Managing Consumer Privacy Concerns in Personalization: A Strategic Analysis of Privacy Protection. MIS Quarterly 2011 Jun;35(2):423-444. [doi: 10.2307/23044050]

Abbreviations

e-commerce: electronic commerce
OHC: online health community

Edited by G Eysenbach; submitted 13.09.19; peer-reviewed by S Atanasova, M Abdelhamid, L Harst, B Seppen; comments to author 07.10.19; revised version received 02.01.20; accepted 25.03.20; published 21.05.20.

Please cite as:
Dang Y, Guo S, Guo X, Vogel D
Privacy Protection in Online Health Communities: Natural Experimental Empirical Study
J Med Internet Res 2020;22(5):e16246
URL: http://www.jmir.org/2020/5/e16246/
doi:10.2196/16246
PMID:32436851

Original Paper

Mental Health–Related Behaviors and Discussions Among Young Adults: Analysis and Classification

Ryan Rivas[1], BS; Moloud Shahbazi[1], PhD; Renee Garett[2], MSW, LCSW; Vagelis Hristidis[1], PhD; Sean Young[3], MS, PhD

[1]Department of Computer Science and Engineering, University of California, Riverside, Riverside, CA, United States

[2]ElevateU, Los Angeles, CA, United States

[3]University of California Institute for Prediction Technology, University of California, Irvine, Irvine, CA, United States

Corresponding Author:
Ryan Rivas, BS
Department of Computer Science and Engineering
University of California, Riverside
363 Winston Chung Hall
900 University Ave
Riverside, CA
United States
Phone: 1 9518272838
Email: rriva002@ucr.edu

Abstract

Background: There have been recurring reports of web-based harassment and abuse among adolescents and young adults through anonymous social networks.

Objective: This study aimed to explore discussions on the popular anonymous social network Yik Yak related to social and mental health messaging behaviors among college students, including cyberbullying, to provide insights into mental health behaviors on college campuses.

Methods: From April 6, 2016, to May 7, 2016, we collected anonymous conversations posted on Yik Yak at 19 universities in 4 different states and performed statistical analyses and text classification experiments on a subset of these messages.

Results: We found that prosocial messages were 5.23 times more prevalent than bullying messages. The frequency of cyberbullying messages was positively associated with messages seeking emotional help. We found significant geographic variation in the frequency of messages offering supportive vs bullying messages. Across campuses, bullying and political discussions were positively associated. We also achieved a balanced accuracy of over 0.75 for most messaging behaviors and topics with a support vector machine classifier.

Conclusions: Our results show that messages containing data about students' mental health–related attitudes and behaviors are prevalent on anonymous social networks, suggesting that these data can be mined for real-time analysis. This information can be used in education and health care services to better engage with students, provide insight into conversations that lead to cyberbullying, and reach out to students who need support.

(J Med Internet Res 2020;22(5):e17224) doi:10.2196/17224

KEYWORDS

social media; data analysis; supervised machine learning; universities; students

Introduction

Background

The transition from high school to college marks the beginning of an important period of psychosocial development. The academic and social demands of college life are often rigorous and can pose a risk to undergraduate students' health and well-being [1]. One example of the challenges they face is poor sleep [2], which has been linked to a number of adverse consequences, including higher rates of depressive symptoms and stress [3,4], weight gain [5], and poor academic performance [6]. Another concern for undergraduate students that has arisen in recent years is their social media use, as studies show a link between cyberbullying and major health problems such as

substance use, depression, poor sleep, and suicide [7-9]. Given the array of health risks faced by undergraduate students, it is important to be aware of students' health and risk-related behaviors to be able to provide adequate services and support, such as from psychological and medical campus services.

Traditionally, methods for monitoring the health of a population, for example, students on a college campus, have focused on case reports and surveys [10,11]. Although these methods can offer insights into health-related attitudes and behaviors, they can be time- and cost-intensive to implement. However, researchers using social media data can collect and analyze behavior data in real time [10,11], allowing health authorities to address student needs in a flexible and timely manner.

To explore the feasibility of using social media platforms to identify and predict health-related events, Young et al [12] screened geolocated Twitter messages for keywords that suggested HIV risk behaviors. The authors used negative binomial regression analyses to determine the association between tweets about HIV risk behaviors and county-level HIV data in the United States. They ran analyses to determine the association between tweets about HIV risk behaviors and county-level HIV data in the United States. The results showed a strong association between tweets about HIV risk behaviors and actual county HIV data. Additionally, De Choudhury et al [13] successfully used tweets to predict the onset of major depressive disorder with 70% accuracy. They selected tweets based on indicators such as linguistic style, use of terms associated with depression, and social network characteristics.

Yik Yak was an anonymous web-based bulletin board for users within the same geographic area (eg, college campuses) that debuted in 2013 [14]. At the time of this study, it was a popular social network for college students but faced substantial criticism. Critics argued, aided by anecdotal evidence relayed through media reports, that anonymous posting encourages harassment and bullying [14-17]. In a recent content analysis of Yik Yak conversations [18], there was no evidence of a pervasive culture of harassment and abuse. However, contradictory to this analysis, researchers have observed derogatory and incendiary comments, arguably racist and sexist messages, and several likely instances of bullying [18]. Furthermore, other research has shown that harassment is prevalent among users of Yik Yak and other anonymous social networks in Bangladesh [19]. Although Yik Yak is now defunct, the rising popularity of anonymous social networks [20] suggests that its data can still provide useful insights.

Study Overview

In this study, we explored two types of messages students made on Yik Yak. The first type consists of posts exhibiting messaging behaviors that can have an impact on students' health in relation to cyberbullying. This includes cyberbullying itself, which has previously been linked to health problems [7-9]. It also includes prosocial messages, which are messages sent by a user with the intention of benefiting one or more other users [21], or with the intention of seeking such messages. The prosocial messaging behaviors we selected are related to bullying and its effects on health. Two of these are seeking and offering support, as students with high depression or anxiety often turn to social media for social support [22]. The second type consists of messages that discuss one of 4 topics frequently discussed by students on Yik Yak, such as relationships and living on campus, to provide additional context to the messaging behaviors we analyzed in this study. We analyze these messaging behaviors and topics by determining which ones are most frequently discussed and which are the most popular (in terms of votes) and by finding correlations between different messaging behaviors and topics.

Our goal is to provide insights for school administrators, public health researchers, and health care professionals regarding the prevalence of messaging behaviors, such as bullying and social support, and knowledge of general topics discussed in the network. Specifically, the purpose of this study is to show that messaging behaviors that can have an impact on students' health occur frequently on anonymous social networks, demonstrate how they are regarded by other students by analyzing their popularity, describe the prevalence and popularity of topics that are commonly discussed by college students, and explore the intercorrelations between these messaging behaviors and topics. Knowledge of these activities on anonymous social networks can inform interventions that promote healthy and prosocial behaviors among adolescents and young adults.

We also investigated the feasibility of automatic classification of messaging behaviors and topics in this study. This involved training 3 machine learning algorithms with several combinations of hyperparameters to determine the best combination for each messaging behavior and topic. We report the results of these models on test data to demonstrate their effectiveness. An accurate classification model can complement the insights provided by this study by providing administrators, researchers, and health care professionals with a tool to more easily find relevant messages.

Methods

Data

From April 6, 2016, to May 7, 2016, we collected anonymous conversations posted on the Yik Yak social network at 5 randomly selected universities located in each of the 4 most populous US states: California (CA); Florida (FL); New York (NY); and Texas (TX). To protect our analyses from the influence of a university with an exceptionally large number of messages, we calculated the number of messages from each university per capita with respect to the number of students enrolled at that university. We then flagged universities that had a number of messages per enrolled student more than 1.5 SDs above their state's mean. This resulted in the removal of 1 university, the University of Texas at Dallas, leaving a total of 19 universities. Table 1 lists these universities, their status as either a public or a private university, their enrollment, and their ranking according to the 2017 *Wall Street Journal*/Times Higher Education College Rankings [23]. Enrollment and rankings are used as part of our analysis of the interplay between variables. For our analysis, we randomly selected 100 conversation threads from each of the universities (N=16,966 messages), with a mean of 892.95 (SD 128) messages per university. We analyzed the messages with respect to the type

of messaging behavior, content, and popularity of message type and content.

Table 1. Characteristics of universities included in the study.

State and university	Public or private	Enrollment	Ranking
CA[a]			
California Polytechnic State University	Public	19,226	221
CSU[b] Chico	Public	16,535	467
CSU Los Angeles	Public	20,353	700
CSU San Bernardino	Public	17,167	700
University of California, Irvine	Public	25,001	153
FL[c]			
Florida International University	Public	53,525	550
Florida State University	Public	36,575	226
University of Central Florida	Public	59,894	445
University of Florida	Public	36,731	56
University of South Florida	Public	35,035	396
NY[d]			
Cornell University	Private	14,706	9
CUNY[e] Hunter College	Public	20,582	350
CUNY John Jay College of Criminal Justice	Public	15,845	700
SUNY[f] Buffalo State	Public	10,665	700
SUNY New Paltz	Public	7756	423
TX[g]			
Tarleton State University	Public	11,008	800
Texas Tech University	Public	29,342	550
University of Houston	Public	36,128	388
University of Texas, Rio Grande Valley	Public	27,560[h]	N/A[i]

[a]CA: California.

[b]CSU: California State University.

[c]FL: Florida.

[d]NY: New York.

[e]CUNY: City University of New York.

[f]SUNY: State University of New York.

[g]TX: Texas.

[h]Fall 2016 enrollment for the University of Texas Rio Grande Valley [24].

[i]N/A: not applicable.

Messaging Behaviors

Within the context of this study, we use the term *messaging behavior* to refer to the intent of a message, that is, what a user is trying to accomplish by posting a message. For each message, we determined if it displayed 1 of the 4 predefined messaging behaviors listed in Table 2. Among these is bullying, which we included in our analysis because of its effects on student health [7-9]. A message was considered to be bullying if it intended harm (ie, if the purpose of the message appeared to be to negatively impact the recipient's mental health), was indicative of a power imbalance (eg, the message was racist or sexist), and if the sender repeatedly sent these messages [25]. We also included seeking help and offering support because of their relation to health and bullying—supportive environments can be seen as more healthy and possibly more likely to prevent or reduce bullying. Humor was included to better understand if users were intentionally bullied or trying to be humorous. A total of 2 undergraduate raters independently coded the selected messages for these 4 messaging behaviors; each message was assigned a messaging behavior only if both raters coded it as such.

Table 3 lists the range, SD, mean, and median for several characteristics of messages with the messaging behaviors defined in Table 2: message length, measured in both characters and words; the number of replies received by any message; the number of replies received by initial posts (ie, the first message in a thread); the post time for messages posted between midnight and noon (AM); and the post time for messages posted between noon and midnight (PM).

Table 2. Definitions of messaging behaviors included in the study.

Messaging behavior	Definition	Examples	Cohen kappa (number of agreements)
Seeking help	Seeking social support (eg, emotional support and help with problems) from other users	• "I like don't know what to do with myself. Literally I have no one to talk to" • "What's the easiest class to fill art requirement? I'm terrible at art"	0.48 (90)
Offering support	Giving social support to other users	• "Hope everything gets resolved OP!" • "You've got this!"	0.56 (86)
Bullying	Intends harm, indicative of a power imbalance, and messages are repeatedly sent [25]	• "You people are disgusting" • "In the words of DJ Khaled 'congratulations you played yourself' it's not hard to portray being a moron. It's quite sad actually"	0.00 (95)
Humor	Intends to be funny without bullying	• "I predict my day based on my morning poo" • "Why get thinner when you can get more dinner?"	0.48 (87)

Table 3. Characteristics of messages with each messaging behavior.

Characteristic	Range	Mean (SD)	Median
Seeking help			
Message length			
Characters	11-204	74.10 (47.82)	61
Words	2-42	14.61 (9.60)	12
Number of replies			
All posts	0-50	4.14 (7.03)	2
Initial post	0-50	5.47 (7.61)	3
Post time			
AM	12:01 AM-11:48 AM	3:38 (2:56)	2:43
PM	12:06 PM-11:57 PM	7:33 (3:13)	8:06
Offering support			
Message length			
Characters	2-200	74.87 (58.39)	58
Words	1-43	14.39 (11.25)	11
Number of replies			
All posts	0-17	0.04 (0.66)	0
Initial post	0-17	4.57 (6.50)	1
Post time			
AM	Midnight-11:57 AM	3:27 (2:43)	2:47
PM	Noon-11:59 PM	7:44 (3:01)	8:25
Bullying			
Message length			
Characters	3-230	63.26 (49.64)	47
Words	1-40	11.92 (9.32)	9
Number of replies			
All posts	0-44	0.17 (2.42)	0
Initial post	0-44	4.07 (11.53)	1
Post time			
AM	Midnight-11:58 AM	3:37 (2:32)	3:12
PM	12:10 PM-11:58 PM	8:38 (3:00)	9:43
Humor			
Message length			
Characters	2-199	32.37 (43.96)	36
Words	1-41	6.37 (8.43)	7
Number of replies			
All posts	0-9	0.28 (1.02)	0
Initial post	0-9	1.83 (2.02)	1
Post time			
AM	12:02 AM-11:58 AM	3:21 (2:49)	2:40
PM	12:09 PM-23:59 PM	7:17 (3:25)	8:09

Message Topics

We applied latent Dirichlet allocation (LDA) to the message corpus to identify themes within the message content. LDA is a common method for categorizing topics and themes [26]. Each topic, in turn, is probabilistically associated with various words. As topics are defined purely in statistical terms, the user chooses its semantic interpretation (ie, its label) based on word probabilities for the topic.

Next, we sought to identify topics in which the LDA message classifications aligned most closely with human judgment. We did this with a subset of 1200 randomly selected messages to which the LDA assigned a topic with a probability greater than 0.7. For each of these messages, a team of 3 raters decided if the LDA topic assignment was correct (ie, does the message discuss topic *X*). On the basis of these results, we selected the 4 topics with the highest classification accuracy: relationships and sex, college living, politics, and school and classes.

In the final step, 2 undergraduate raters independently applied the 4-topic classification scheme to 96 randomly selected messages. We found that their interrater agreement was high (Cohen kappa=0.78), so all remaining messages were coded by 1 of the 2 raters. Table 4 lists Cohen kappa for each individual topic; it is undefined for politics because neither rater coded any of the 96 messages for that topic.

Table 5 lists the range, SD, mean, and median for several characteristics of messages with these topics.

Table 4. Cohen kappa for each topic (n=96).

Statistic	Relationships and sex	College living	Politics	School and classes
Cohen kappa	0.73	1.00	Undefined	0.77
Number of agreements	90	96	96	91

Table 5. Characteristics of messages with each topic.

Characteristic	Range	Mean (SD)	Median
Relationships and sex			
Message length			
Characters	2-252	82.18 (52.32)	70
Words	1-47	16.17 (10.32)	14
Number of replies			
All posts	0-50	0.96 (3.43)	0
Initial post	0-50	4.60 (6.31)	3
Post time			
AM	Midnight-11:58 AM	3:27 (2:21)	3:07
PM	Noon-11:59 PM	8:05 (3:16)	8:55
College living			
Message length			
Characters	3-200	74.56 (49.98)	62
Words	1-42	14.36 (9.52)	12
Number of replies			
All posts	0-19	0.83 (2.15)	0
Initial post	0-19	2.60 (3.14)	2
Post time			
AM	Midnight-11:56 AM	3:34 (2:38)	2:57
PM	Noon-11:59 PM	6:57 (3:15)	7:24
Politics			
Message length			
Characters	5-210	107.72 (58.43)	99
Words	1-43	19.22 (10.65)	17
Number of replies			
All posts	0-53	0.83 (4.27)	0
Initial post	0-53	7.13 (10.59)	4
Post time			
AM	Midnight-11:47 AM	3:26 (2:32)	3:06
PM	12:08 PM-11:58 PM	7:52 (3:11)	7:30
School and classes			
Message length			
Characters	3-202	71.41 (49.59)	59
Words	1-42	13.67 (9.38)	11
Number of replies			
All posts	0-44	0.98 (3.33)	0
Initial post	0-44	4.39 (5.90)	3
Post time			
AM	Midnight-11:58 AM	3:41 (2:58)	2:46
PM	12:03 PM-11:59 PM	6:58 (3:09)	7:35

Analysis

Our analysis consisted of 3 parts: frequency of messaging behaviors and topics, popularity of messaging behaviors and topics, and interplay between variables. In the first 2 parts, we used messages that raters uniquely assigned to 1 or none of the 4 predefined messaging behaviors to assess the frequency and popularity of messaging behaviors. Similarly, we used messages that raters uniquely assigned to 1 or none of the 4 LDA-derived topics to assess the frequency and popularity of messaging behaviors. In all statistical analyses, the significance criterion was alpha=.05.

In our analysis of the relative frequencies of messaging behaviors and topics on Yik Yak, Bonferroni-corrected Fisher exact tests determined if differences in the frequencies of these messaging behaviors or topics across states were statistically significant. If we found that the differences for a messaging behavior or topic were significant, we followed this up with Bonferroni-corrected Fisher exact tests for pairwise comparisons between states of the frequency of that messaging behavior or topic.

We determined the popularity of a message by the aggregate score of +1 votes (upvotes) and −1 votes (downvotes) assigned by Yik Yak users before data collection. Notably, if a message on Yik Yak reaches a sum score of −5, it is automatically deleted from the social network. Thus, the lowest possible popularity score for a message in our dataset was −4. To protect our analyses from the influence of a few massively popular messages, we flagged messages with a score greater than 2.5 SDs above the grand mean. We then submitted the individual message scores to state × messaging behavior and state × topic analysis of variance (ANOVA), followed up by Tukey range test to further investigate any significant main effects of each ANOVA.

The third part of our analysis examined the relationship between the frequency of prosocial messages in which users sought help or offered support, the frequency of bullying messages, the popularity of these messaging behaviors, and the frequency of topics. We carried out an analysis at the university level. For each university, we calculated mean messaging behavior frequencies, the corresponding mean popularity scores, and mean topic frequencies. We measured correlations between these variables together with 2 additional variables—the number of students enrolled and school ranking.

Classification

We conducted a series of experiments with 3 text classification algorithms on the messaging behaviors and topics in this study. The first 2 are random forest [27] and linear support vector machine (SVM) [28] classifiers with term frequency-inverse document frequency (TF-IDF) vectors [29], and the third is a convolutional neural network (CNN) text classifier [30] with global vectors for word representation (GloVe) [31].

In each experiment, we selected 1 messaging behavior or topic and regarded each message in the dataset as a tuple (t, c), where t is the message text concatenated with tokens for the university and state the message is from, and c is a class label *positive* (the selected messaging behavior or topic is present in the message) or *negative* (the messaging behavior or topic is not present). We randomly selected 10.00% (1697/16,966) of the dataset to be used as the test dataset. With the remaining training dataset, we used 5-fold cross-validation and measured the balanced accuracy [32] of each classifier to determine the best combination of classifier hyperparameters, which are then used with the full training dataset to build the final classifier model.

Table 6 lists the hyperparameters and their respective values evaluated by our experiments for each classifier. For all classifiers, we preprocess the data by removing stop words and lemmatizing the remaining words with the natural language toolkit [33]. For the random forest and SVM classifiers, we add *balanced* class weights as defined by Scikit-learn [34]. The TF-IDF vectors are also built from the implementation in Scikit-learn [34]. The remaining hyperparameters are set to their default values, as defined by the implementations of these classifiers in Scikit-learn [34]. For the CNN classifier, we perform upsampling such that the positive messages in the training data are as frequent as the negative messages and use 100-dimension GloVe vectors pretrained on Twitter data. All other CNN hyperparameters are set to their default values as defined in the code by Ng [35].

Table 6. Classifier hyperparameter values evaluated in our experiments.

Classifier and hyperparameter	Values
Random forest	
Maximum tree depth	2, 4, 8, 16, 32, 64
Number of trees	10, 100, 1000
SVM[a]	
C[b]	0.001, 0.01, 0.1, 1, 10
Loss function	Hinge, squared hinge
CNN[c]	
Filter window sizes	(2, 3, 4), (3, 4, 5), (4, 5, 6)
Feature maps per filter window size	100, 200, 300, 400, 500, 600

[a]SVM: support vector machine.

[b]*C*: SVM regularization parameter.

[c]CNN: convolutional neural network.

Results

Frequency of Messaging Behaviors

A total of 11.91% (2021/16,966) of the messages were focused on 1 of the 4 predefined messaging behavior categories: seeking help, offering support, humor, and bullying. Table 7 lists the frequencies of these messaging behaviors by state. We found significant differences in the relative frequency of messages offering support ($P<.001$) and bullying messages ($P<.001$). We found no significant geographic differences for messages seeking help ($P=.20$) or for humorous messages ($P=.40$). Using separate Fisher exact tests, we found that the 2 states with the lowest rates of bullying, CA and FL, differed significantly from the states with the highest rates, NY and TX ($P<.001$ for CA vs TX and FL vs TX, $P=.001$ for CA vs NY, $P=.003$ for FL vs NY).

Table 7. Frequency of messaging behaviors by state.

Messaging behavior	CA[a] (N=4496), n (%)	FL[b] (N=4694), n (%)	NY[c] (N=4273), n (%)	TX[d] (N=3503), n (%)	Total (N=16,966), n (%)	Bonferroni-corrected Fisher exact *P* value
Seeking help	70 (1.56)	94 (2.00)	65 (1.52)	70 (2.00)	299 (1.76)	.20
Offering support	183 (4.07)	381 (8.12)	234 (5.48)	88 (2.51)	886 (5.22)	<.001
Bullying	61 (1.36)	68 (1.45)	98 (2.29)	93 (2.65)	320 (1.96)	<.001
Humor	140 (3.11)	134 (2.85)	144 (3.37)	98 (2.80)	516 (3.15)	.40

[a]CA: California.

[b]FL: Florida.

[c]NY: New York.

[d]TX: Texas.

We also evaluated a sample of messages that were not assigned any of the 4 predefined messaging behavior categories to better understand the nature of messaging behavior outside of these categories. This sample consisted of 100 messages that were the first messages in their respective conversation threads. We found that the majority of these messages (68/100) were commentary, for example, anticipation of future events (“Cant wait for summer!!! #summer16”), reactions to personal experiences (“I hate when people tell me to put on headphones.”), and observations (“So many economics majors on yikyak nowadays”). Other messages (16/100) asked questions that did not seek social support, for example, soliciting opinions (“Do you think all pedophiles should be executed or do you think they deserve a 2nd chance and then should be executed if they relapse?”) and polling (“Quick poll. What's your ethnicity?”). Further messages (12/100) sought people to meet with or talk to for purposes other than social support, for example, for dating (“Any cute girls in the dorms? Drop your snapchat names”) or classes (“Anyone in geology 210 on M for 4:00-5:50?”).

The remaining messages in the sample (4/100) lacked sufficient context to judge their messaging behavior. Although these broadly defined messaging behaviors are not directly related to this study and, thus, not subjected to further analysis, this sample of posts shows that future work focusing on the commentary present on an anonymous social network would likely have substantial coverage of the message content of that network.

Frequency of Topics

Using only messages with 1 or none of the 4 LDA-derived topics (relationships and sex, college living, politics, and school and classes), we excluded 0.69% (117/16,966) of the messages from

the frequency analysis. A total of 26.33% (4437/16,849) of the remaining messages dealt with either relationships and sex (2516/16,849, 14.93%), college living (644/16,849, 3.82%), politics (607/16,849, 3.60%), or school and classes (670/16,849, 3.98%). In Table 8, we break these numbers down further by state. Using separate Fisher exact tests, we found significant regional differences for each topic. NY had the fewest relationship messages and differed significantly from CA (P<.001) and TX (P=.048).

Table 8. Frequency of topics by state.

Topics	CA[a] (N=4443), n (%)	FL[b] (N=4668), n (%)	NY[c] (N=4253), n (%)	TX[d] (N=3485), n (%)	Total (N=16,849), n (%)	Bonferroni-corrected Fisher exact *P* value
Relationships and sex	730 (16.43)	689 (14.76)	532 (13.21)	535 (15.35)	2516 (14.93)	<.001
College living	224 (5.04)	83 (1.78)	157 (3.69)	180 (5.16)	644 (3.82)	<.001
Politics	133 (2.99)	122 (2.61)	317 (7.45)	35 (1.00)	607 (3.60)	<.001
School and classes	208 (4.68)	114 (2.44)	150 (3.53)	198 (5.68)	670 (3.98)	<.001

[a]CA: California.

[b]FL: Florida.

[c]NY: New York.

[d]TX: Texas.

We followed up on these significant effects with Bonferroni-corrected Fisher exact tests for all pairwise comparisons between states for each topic. We found significant differences in the number of college living messages between all states (P<.001), except for CA and TX, the 2 states with the most college living messages (P=.76). Finally, we found significant differences in the frequency of school-related messages between states (P<.001); CA and TX, where school was discussed the most, had the least significant difference (P=.04).

Popularity of Messaging Behaviors

In this and the following section, we report findings on the popularity of the different messaging behaviors and topics, based on the aggregate of +1 votes (upvotes) and −1 votes (downvotes) each message elicited from Yik Yak users. We identified 1.80% (305/16,966) of the messages as popularity outliers and excluded these from further analysis.

Table 9 displays the mean popularity scores for the 4 messaging behaviors (seeking help, offering support, bullying, and humor) at the state level (CA, FL, NY, and TX). We submitted the individual message scores to a state × messaging behavior ANOVA. Both main effects were significant: $F_{3,1940}$=5.11, mean square error (MSE)=4.1, and P=.002 for state and $F_{3,1940}$=25.85, MSE=4.1, and P<.001 for messaging behavior. The interaction between the 2 factors was not significant ($F_{9,1940}$=0.94; MSE=4.1; P=.49).

Table 9. Popularity of messaging behaviors and topics by state.

Messaging behavior	CA[a]		FL[b]		NY[c]		TX[d]		Total	
	Mean[e] (SE)	n	Mean (SE)	n	Mean (SE)	n	Mean (SE)	n	Mean (SE)	n
Seeking help	1.04 (0.26)	68	1.37 (0.21)	92	0.78 (0.30)	63	0.53 (0.27)	70	0.97 (0.13)	293
Offering support	1.00 (0.11)	182	0.98 (0.08)	380	1.22 (0.12)	230	0.77 (0.16)	88	1.03 (0.06)	880
Bullying	0.40 (0.32)	58	0.32 (0.17)	68	0.59 (0.23)	96	0.32 (0.18)	92	0.42 (0.11)	314
Humor	1.50 (0.20)	124	1.71 (0.22)	125	2.14 (0.27)	130	1.27 (0.20)	90	1.69 (0.12)	469

[a]CA: California.

[b]FL: Florida.

[c]NY: New York.

[d]TX: Texas.

[e]Mean: Mean message popularity scores are based on the aggregate number of upvotes (+1) and downvotes (−1) per message.

We used Tukey range test to determine which state exhibited significantly different mean popularity scores. This analysis revealed that, on average, Yik Yak messages received lower popularity scores in TX than in FL (P=.03) and NY (P<.001). Additionally, Tukey test showed that bullying messages were the least popular and differed significantly from messages seeking help (P=.003), messages offering support (P<.001), or humorous messages (P=.001). In contrast, humorous messages were the most popular and scored significantly higher than the other 3 message types (all P<.001).

Popularity of Topics

Table 10 summarizes the mean popularity scores of messages that discussed 1 of the 4 topics identified through LDA: relationships and sex, college living, politics, or school and classes. A state (CA, FL, NY, and TX) × topic ANOVA revealed main effects of $F_{3,4293}$=11.23, MSE=4.9, and P<.001 for state

and $F_{3,4293}$=7.32, MSE=4.9, and P<.001 for the topic as well as a significant state-by-topic interaction of $F_{9,4293}$=2.52, MSE=4.9, and P=.007. We carried out Tukey test to further investigate the significant main effects. We found that TX, the state with the lowest popularity scores overall, differed significantly from CA (P<.001), FL (P=.03), and NY (P<.001). Regarding the popularity of topics, school and classes was a significantly less popular topic than relationships and sex (P=.002), college living (P=.002), and politics (P=.001).

Table 10. Popularity of topics by state.

Topic	CA[a]		FL[b]		NY[c]		TX[d]		Total	
	Mean[e] (SE)	n	Mean (SE)	n	Mean (SE)	n	Mean (SE)	n	Mean (SE)	n
Relationships and sex	1.56 (0.09)	700	1.03 (0.08)	678	1.16 (0.10)	548	0.96 (0.08)	528	1.19 (0.05)	2454
College living	1.31 (0.15)	209	1.56 (0.26)	78	1.70 (0.23)	146	0.78 (0.14)	175	1.28 (0.09)	608
Politics	1.17 (0.21)	129	1.46 (0.24)	119	1.34 (0.14)	314	1.49 (0.43)	35	1.34 (0.10)	597
School and classes	0.84 (0.12)	197	1.09 (0.20)	114	1.08 (0.18)	145	0.43 (0.09)	194	0.82 (0.07)	650

[a]CA: California.

[b]FL: Florida.

[c]NY: New York.

[d]TX: Texas.

[e]Mean: Mean message popularity scores are based on the aggregate number of upvotes (+1) and downvotes (−1) per message.

The significant state-by-topic interaction indicates that states differ with respect to the relative popularity of topics. To identify patterns of topic popularity within each state, we conducted ANOVAs with topic as a single factor, separately for each state. These ANOVAs yielded a significant effect of topic for CA ($F_{3,1231}$=5.36; MSE=5.39; P=.001) and TX ($F_{3,928}$=5.84; MSE=3.17; P<.001) but not for FL ($F_{3,985}$=2.41; MSE=4.91; P=.07) or NY ($F_{3,\ 1149}$=2.34; MSE=5.7; P=.07). We followed up on the significant effects for CA and TX using Tukey test. In CA, school and classes were a less popular topic than relationships and sex (P<.001). In TX, messages about school and classes were less popular than messages about relationships (P=.002) and politics (P<.009).

Interplay Between Variables

We summarize the intercorrelations between the frequency of prosocial messages in which users sought help or offered support, the frequency of bullying messages, the popularity of these messaging behaviors, the frequency of topics, and school enrollment and ranking in Table 11. These correlations are based on 19 schools, except for correlations involving the variable *ranking*, for which n=18.

We found that schools with a greater frequency of help-seeking messages also exhibited a greater frequency of messages offering support (P=.04). Campuses where students posted less about relationships and sex sent more messages offering support (P=.002). Moreover, messages offering support were more frequent at higher-ranking schools (P=.006). Bullying occurred more often on campuses where users posted more about politics (P=.048) and where messages seeking help were popular (P=.02). Messages offering support were more popular at campuses where students posted more about classes (P=.04). Finally, we found that the frequency of messages about college living was positively related to the frequency of messages about classes (P=.04) but negatively related to the number of enrolled students (P=.05). The remaining correlations in Table 9 were not statistically significant.

Table 11. Intercorrelations at the school level.

Variable	SH[a]	OS[b]	BU[c]	PH[d]	PS[e]	PB[f]	RS[g]	CL[h]	PO[i]	SC[j]	EN[k]	RA[l]
SH	__[m]	0.48	−0.13	−0.06	0.37	0.01	−0.35	0.01	−0.38	0.36	0.17	−0.29
OS	__[n]	__[m]	−0.33	0.16	0.00	0.05	−0.66	−0.30	0.07	−0.08	0.20	−0.62
BU	__[n]	__[n]	__[m]	0.52	0.37	−0.35	0.36	0.01	0.46	−0.07	−0.07	0.10
PH	__[n]	__[n]	__[n]	__[m]	0.37	−0.02	0.19	−0.03	0.30	−0.11	0.90	−0.21
PS	__[n]	__[n]	__[n]	__[n]	__[m]	−0.18	0.26	0.19	0.16	0.47	−0.15	−0.17
PB	__[n]	__[n]	__[n]	__[n]	__[n]	__[m]	−0.20	−0.11	0.13	0.03	−0.21	−0.08
RS	__[n]	__[n]	__[n]	__[n]	__[n]	__[n]	__[m]	0.09	−0.09	−0.02	0.09	0.29
CL	__[n]	__[n]	__[n]	__[n]	__[n]	__[n]	__[n]	__[m]	−0.14	0.47	−0.45	0.29
PO	__[n]	__[n]	__[n]	__[n]	__[n]	__[n]	__[n]	__[n]	__[m]	−0.19	−0.27	−0.35
SC	__[n]	__[n]	__[n]	__[n]	__[n]	__[n]	__[n]	__[n]	__[n]	__[m]	−0.26	−0.01
EN	__[n]	__[n]	__[n]	__[n]	__[n]	__[n]	__[n]	__[n]	__[n]	__[n]	__[m]	−0.33
RA	__[n]	__[n]	__[n]	__[n]	__[n]	__[n]	__[n]	__[n]	__[n]	__[n]	__[n]	__[m]

[a]SH: seeking help.

[b]OS: offering support.

[c]BU: bullying.

[d]PH: popularity of seeking help.

[e]PS: popularity of offering support.

[f]PB: popularity of bullying.

[g]RS: relationships and sex.

[h]CL: college living.

[i]PO: politics.

[j]SC: school and classes.

[k]EN: enrollment.

[l]RA: ranking.

[m]Cells along the diagonal represent the same variable in both row and column, thus no correlation is reported.

[n]Cells below the diagonal duplicate those above the diagonal and are left blank for clarity.

Classification Results

Tables 12 and 13 summarize the results of our trained classifiers on the test data. As accuracy can be misleadingly high for imbalanced datasets, we also report balanced accuracy. Using this metric, we see that SVM has the best performance on 5 messaging behaviors and topics (offering support, bullying, relationships and sex, politics, and school and classes), with a balanced accuracy of over 0.75 on all but the humor dataset and an average balanced accuracy of 0.7827. CNN was the second-best performer, with the best performance on humor and college living and an average balanced accuracy of 0.7645.

Table 12. Messaging behavior classification results.

Metric and classifier	Seeking help	Offering support	Bullying	Humor
Accuracy				
Random forest	*0.9269* [a]	*0.8120*	*0.9299*	0.6417
SVM[b]	0.6771	0.7501	0.9240	*0.8385*
CNN[c]	0.9098	0.6618	0.9146	0.7195
Balanced accuracy				
Random forest	*0.8575*	0.7151	0.6763	0.6392
SVM	0.8007	*0.7514*	*0.7750*	0.6543
CNN	0.6557	0.7313	0.7702	*0.6942*

[a]The highest accuracy and balanced accuracy achieved for each messaging behavior are italicized for emphasis.

[b]SVM: support vector machine.

[c]CNN: convolutional neural network.

Table 13. Topic classification results.

Metric and classifier	Relationships and sex	College living	Politics	School and classes
Accuracy				
Random forest	0.8209	*0.9028* [a]	0.8704	0.9387
SVM[b]	*0.8521*	0.8981	*0.9405*	*0.9499*
CNN[c]	0.7943	0.8533	0.9399	0.9010
Balanced accuracy				
Random forest	0.7380	0.7323	0.7775	0.7899
SVM	*0.8145*	0.7842	*0.8605*	*0.8212*
CNN	0.7902	*0.8075*	0.8524	0.8147

[a]The highest accuracy and balanced accuracy achieved for each topic are italicized for emphasis.

[b]SVM: support vector machine.

[c]CNN: convolutional neural network.

Discussion

Principal Findings

Owing to the growing popularity of social media across all segments of society, researchers have a plethora of data sources from which they can derive new insights about people's social and health-related attitudes, behaviors, and beliefs. The ability to observe social media users in near real time holds particular promise in the domain of public health and health care, where rapid detection of health-relevant events and timely intervention are essential. This study aimed to explore the prevalence of information pertaining to college students' health and well-being contained in their conversations on an anonymous social network. To this end, we analyzed the frequency and popularity of prosocial messages and bullying messages as well as the frequency and popularity of topics discussed on the web.

In our dataset, prosocial messages (seeking help, offering support, and humor) appeared more frequently than bullying messages (1735/16,966, 10.23% vs 332/16,966, 1.96%), and there were significant regional differences in the frequency of messages associated with support or bullying. Notably, Yik Yak users attending TX colleges sent the fewest supportive messages and the most bullying messages. We should interpret this finding with caution in light of the relatively small number of messages and universities considered for our study. Nevertheless, this finding highlights a potentially problematic pattern of social media use among college students that future research may link to adverse health outcomes. Unsurprisingly, bullying messages were the least popular, and humorous messages were the most popular among Yik Yak users, independent of the state in which they lived.

To identify the topics of Yik Yak messages, we relied on statistical modeling as an alternative to the subjective classification scheme recently used by Black et al [18]. A subsequent analysis of topic prevalence revealed that relationships and sex was the most frequently discussed topic among college students. School and classes turned out to be the least popular topic, as measured by the number of upvotes and downvotes a message received. From an intervention point of view, regional differences in topic frequency and popularity matter because they offer campus representatives and health

professionals clues on how to best engage a student population, both on the web and offline. Although the relative popularity of topics was similar across states, we found greater regional variation in the relative frequency of topics. For example, 7.44% (318/4273) of Yik Yak messages in the state of NY discussed politics compared with only 1.00% (35/3503) in TX, and college living was addressed in 5.60% (252/4496) of messages in CA but in only 2.28% (107/4694) of messages in FL.

With our final correlational analysis, we wanted to learn more about factors that promote prosocial web-based behaviors and prevent cyberbullying at US colleges. Several findings are worth noting. At schools where students often sought help through messages, messages offering support were also more frequent. We speculate that students may offer support in response to requests for help, but the reverse relationship is also conceivable: at schools where support is offered frequently, students may feel encouraged to ask for help. A higher prevalence of supportive messages also appears to be a characteristic of higher-ranking universities. Although the *Wall Street Journal*/Times Higher Education's college rankings [23] do not take into account social support between students, some hidden factors that lead to a higher prevalence of social support may have also been indirectly captured by their methodology. Our observation of a positive relationship between the popularity of messages offering support and the frequency of the school and classes topic may be explained by a positive response, in the form of upvotes, to support offered to students expressing frustrations with coursework and exams. It is more difficult to interpret why messages of support were sent more often at schools where relationships and sex were discussed less frequently. This requires further investigation.

Two results speak directly to the frequency of cyberbullying on college campuses. First, there was a positive relationship between bullying and the popularity of messages seeking help. One interpretation for this finding is that students react prosocially to a higher prevalence of bullying by encouraging help-seeking behavior, although they did not appear to actually offer more support (the correlation between the frequency of supporting and bullying messages was negative and not significant). An alternative hypothesis is that certain prosocial messaging behaviors can trigger cyberbullying. Additionally, students at schools with a higher incidence of bullying frequently discussed politics. This result is unsurprising given the often-heated nature of political discussions.

Of the results regarding the frequency of messages about college living, the positive relationship with the frequency of messages about classes is understandable, given that these 2 topics reflect much of the college experience. However, messages about college living are less frequent at schools with lower enrollment rates. One possible explanation may be that smaller schools have less on-campus housing relative to the number of students, but further study is necessary to make this determination.

Our text classification experiments demonstrate the feasibility of automatic classification of the messaging behaviors and topics in this study. The balanced accuracy of the SVM classifier on the test data was reasonably high for most messaging behaviors and topics. Its worst performance was with the humor dataset, which also had the lowest balanced accuracy with the random forest classifier and the second lowest balanced accuracy with the CNN classifier. This may be because of the complexity of humor—forms of humor such as innuendo, sarcasm, and satire may be difficult for a machine learning algorithm to identify.

Conclusions

This study has strong implications for education, public health, and broader fields of health care. Educators could use similar methods to find topics that may be engaging to students on campus. In particular, campus administrators and health service units could identify topic areas where students could engage in a campus-wide dialogue. This could also be helpful for public health professionals because it would provide insight into campus conversations that lead to bullying or hostility. Educators and clinicians could work together to foster a healthier dialogue around the subject and encourage a campus culture of reaching out to fellow students to offer support. In addition to gaining insights into conversations on college campuses, this study represents a first step in guiding research focused on anonymous social networks. The results of this study can help promote the labeling and mining of social data to help students, parents, administrators, and health care workers identify cyberbullying and design interventions to stop it.

This type of work naturally presents opportunities for computer scientists working in health services as well. Mining data from anonymous social networks can extend beyond the college campus and to the public. Computer scientists can design tools to mine and categorize public social data and help create an even farther-reaching monitoring system for educators and public health professionals [36].

The major limitations of this study include the small number of colleges and universities considered, the lack of ability to generalize as Yik Yak has closed down since this study was conducted, the modest number of Yik Yak messages per school, and the relatively small number of classifier hyperparameters evaluated. We, therefore, caution against generalizing our findings until they can be replicated with larger samples and on other anonymous social networks. The main intention of this study was to understand students' web-based behaviors and interests from their messages on an anonymous social network and, more specifically, to garner initial insight into conditions affecting prosocial and antisocial uses of social media that could be integrated into health services. We believe that the findings reported here can be a stepping stone to further research on this topic as well as differences in health behaviors and risks communicated on anonymous social networks vs nonanonymous social networks.

Authors' Contributions

RR performed text classification experiments, determined statistical characteristics of messaging behaviors and topics, and assisted in writing the manuscript. MS collected and analyzed the data and wrote part of the manuscript. RG worked on the analysis and writing. VH led the technical research. SY led the problem formulation and discussion.

Conflicts of Interest

None declared.

References

1. Garett R, Liu S, Young SD. A longitudinal analysis of stress among incoming college freshmen. J Am Coll Health 2017 Jul;65(5):331-338 [FREE Full text] [doi: 10.1080/07448481.2017.1312413] [Medline: 28362146]
2. Vail-Smith K, Felts WM, Becker C. Relationship between sleep quality and health risk behaviors in undergraduate college students. Coll Stud J 2009 Sep 1;43(3):924-931 [FREE Full text]
3. Galambos NL, Dalton AL, Maggs JL. Losing sleep over it: daily variation in sleep quantity and quality in Canadian students. J Res Adolesc 2009 Dec;19(4):741-761. [doi: 10.1111/j.1532-7795.2009.00618.x]
4. Williams PG, Moroz TL. Personality vulnerability to stress-related sleep disruption: pathways to adverse mental and physical health outcomes. Pers Individ Differ 2009 Apr 1;46(5-6):598-603. [doi: 10.1016/j.paid.2008.12.017]
5. Roane BM, Seifer R, Sharkey KM, van Reen E, Bond TL, Raffray T, et al. What role does sleep play in weight gain in the first semester of university? Behav Sleep Med 2015 Nov 2;13(6):491-505 [FREE Full text] [doi: 10.1080/15402002.2014.940109] [Medline: 25115969]
6. Curcio G, Ferrara M, de Gennaro L. Sleep loss, learning capacity and academic performance. Sleep Med Rev 2006 Oct 1;10(5):323-337. [doi: 10.1016/j.smrv.2005.11.001] [Medline: 16564189]
7. Bauman S, Toomey RB, Walker JL. Associations among bullying, cyberbullying, and suicide in high school students. J Adolesc 2013 Apr 1;36(2):341-350. [doi: 10.1016/j.adolescence.2012.12.001] [Medline: 23332116]
8. Gámez-Guadix M, Orue I, Smith PK, Calvete E. Longitudinal and reciprocal relations of cyberbullying with depression, substance use, and problematic internet use among adolescents. J Adolesc Health 2013 Oct 1;53(4):446-452. [doi: 10.1016/j.jadohealth.2013.03.030] [Medline: 23721758]
9. Wang J, Nansel TR, Iannotti RJ. Cyber and traditional bullying: differential association with depression. J Adolesc Health 2011 Apr 1;48(4):415-417 [FREE Full text] [doi: 10.1016/j.jadohealth.2010.07.012] [Medline: 21402273]
10. Young SD. Behavioral insights on big data: using social media for predicting biomedical outcomes. Trends Microbiol 2014 Nov 1;22(11):601-602 [FREE Full text] [doi: 10.1016/j.tim.2014.08.004] [Medline: 25438614]
11. Young SD. A 'big data' approach to HIV epidemiology and prevention. Prev Med 2015 Jan;70:17-18 [FREE Full text] [doi: 10.1016/j.ypmed.2014.11.002] [Medline: 25449693]
12. Young SD, Rivers C, Lewis B. Methods of using real-time social media technologies for detection and remote monitoring of HIV outcomes. Prev Med 2014 Jun 1;63:112-115 [FREE Full text] [doi: 10.1016/j.ypmed.2014.01.024] [Medline: 24513169]
13. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting Depression via Social Media. In: Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media. 2013 Presented at: AAAI'13; July 8-11, 2013; Cambridge, MA p. 128-137 URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/view/6124/6351
14. Mahler J. The New York Times. 2015 Mar 9. Who Spewed That Abuse? Anonymous Yik Yak App Isn't Telling URL: http://www.nytimes.com/2015/03/09/technology/popular-yik-yak-app-confers-anonymity-and-delivers-abuse.html [accessed 2016-11-21]
15. Shontell A. Business Insider. 2015 Mar 28. Why a Girl Who Was Viciously Bullied on Yik Yak Now Believes in the Anonymous App's Future URL: http://www.businessinsider.com/elizabeth-long-was-bullied-on-yik-yak-2015-3 [accessed 2016-11-22]
16. Valencia N. CNN. 2014 Mar 7. Yik Yak Chat App Stirring Up Trouble in High Schools URL: https://www.cnn.com/2014/03/07/tech/yik-yak-app-high-school-problems/ [accessed 2016-11-14]
17. Wang G, Wang B, Wang T, Nika A, Zheng H, Zhao BY. Whispers in the Dark: Analysis of an Anonymous Social Network. In: Proceedings of the 2014 Conference on Internet Measurement Conference. 2014 Presented at: IMC'14; November 5-7, 2014; Vancouver, Canada. [doi: 10.1145/2663716.2663728]
18. Black EW, Mezzina K, Thompson LA. Anonymous social media–understanding the content and context of Yik Yak. Comput Hum Behav 2016 Apr 1;57:17-22. [doi: 10.1016/j.chb.2015.11.043]
19. Nova FF, Rifat MR, Saha P, Ahmed SI, Guha S. Online Sexual Harassment Over Anonymous Social Media in Bangladesh. In: Proceedings of the Tenth International Conference on Information and Communication Technologies and Development. 2019 Presented at: ICTD'19; January 4-7, 2019; Ahmedabad, India. [doi: 10.1145/3287098.3287107]
20. Gerhart N, Koohikamali M. Social network migration and anonymity expectations: what anonymous social network apps offer. Comput Hum Behav 2019 Jun 1;95:101-113. [doi: 10.1016/j.chb.2019.01.030]

21. Batson CD, Powell AA. Altruism and prosocial behavior. In: Handbook of Psychology. Hoboken, NJ: John Wiley & Sons; 2003:463-484.
22. Drouin M, Reining L, Flanagan M, Carpenter M, Toscos T. College students in distress: can social media be a source of social support? Coll Stud J 2018 Dec 1;52(4):494-504 [FREE Full text]
23. Times Higher Education. 2016. Wall Street Journal/Times Higher Education College Rankings 2017 URL: http://www.timeshighereducation.com/rankings/united-states/2017 [accessed 2016-12-09]
24. The University of Texas Rio Grande Valley. 2016. Office of Strategic Analysis and Institutional Reporting (SAIR) URL: http://www.utrgv.edu/sair/ [accessed 2016-12-09]
25. Garett R, Lord LR, Young SD. Associations between social media and cyberbullying: a review of the literature. Mhealth 2016 Dec 19;2:46. [doi: 10.21037/mhealth.2016.12.01] [Medline: 28293616]
26. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res 2003 Jan;3:993-1022 [FREE Full text]
27. Breiman L. Random forests. Mach Learn 2001 Oct 1;45(1):5-32 [FREE Full text] [doi: 10.1023/A:1010933404324]
28. Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995 Sep;20(3):273-297 [FREE Full text] [doi: 10.1007/bf00994018]
29. Manning CD, Raghavan P, Schütze H. Scoring, term weighting and the vector space model. In: Introduction to Information Retrieval. Cambridge, England: Cambridge University Press; 2008:109-133.
30. Kim Y. Convolutional Neural Networks for Sentence Classification. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processin. 2014 Presented at: EMNLP'14; October 25-29, 2014; Doha, Qatar p. 1746-1751. [doi: 10.3115/v1/d14-1181]
31. Pennington J, Socher R, Manning CD. Glove: Global Vectors for Word Representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. 2014 Presented at: EMNLP'14; October 25-29, 2014; Doha, Qatar p. 1532-1543. [doi: 10.3115/v1/d14-1162]
32. Brodersen KH, Ong C, Stephan KE, Buhmann JM. The Balanced Accuracy and Its Posterior Distribution. In: Proceedings of the 20th International Conference on Pattern Recognition. 2010 Presented at: ICPR'10; August 23-26, 2010; Istanbul, Turkey. [doi: 10.1109/icpr.2010.764]
33. Bird S, Klein E, Loper E. Natural Language Processing with Python. Sebastopol, CA: O'reilly Media; 2020.
34. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O. Scikit-learn: machine learning in Python. J Mach Learn Res 2011 Oct 1;12(85):2825-2830 [FREE Full text]
35. Ng S. GitHub. 2018 Dec 27. CNNs for Sentence Classification in PyTorch URL: https://github.com/Shawn1993/cnn-text-classification-pytorch [accessed 2019-03-18]
36. Benbow N, Kirkpatrick C, Gupta A, Villamar J, Chernyshov K, Cramer N, et al. An iterative process of integrating and developing big data modeling and visualization tools in collaboration with public health officials. In: Sage Research Methods Cases: Medicine and Health. Thousand Oaks, CA: Sage Publications; 2020.

Abbreviations

ANOVA: analysis of variance
CA: California
CNN: convolutional neural network
FL: Florida
GloVe: Global vectors for word representation
LDA: latent Dirichlet allocation
MSE: mean square error
NY: New York
SVM: support vector machine
TF-IDF: term frequency-inverse document frequency
TX: Texas

Edited by G Eysenbach; submitted 26.11.19; peer-reviewed by M Wick, B Kerr; comments to author 19.01.20; revised version received 01.03.20; accepted 23.03.20; published 29.05.20.

Please cite as:
Rivas R, Shahbazi M, Garett R, Hristidis V, Young S
Mental Health–Related Behaviors and Discussions Among Young Adults: Analysis and Classification
J Med Internet Res 2020;22(5):e17224
URL: http://www.jmir.org/2020/5/e17224/
doi:10.2196/17224
PMID:32469317

Original Paper

Multidimensional Feature Classification of the Health Information Needs of Patients With Hypertension in an Online Health Community Through Analysis of 1000 Patient Question Records: Observational Study

Aijing Luo[1,2], PhD; Zirui Xin[1,2,3], PhD; Yifeng Yuan[1,2,3], MS; Tingxiao Wen[3], PhD; Wenzhao Xie[1,2], PhD; Zhuqing Zhong[1,2], PhD; Xiaoqing Peng[1,2], PhD; Wei Ouyang[1,2,3], PhD; Chao Hu[2,4,5], PhD; Fei Liu[1,2,3], MS; Yang Chen[1,2,3], MS; Haiyan He[1,2,3], PhD

[1]The Third Xiangya Hospital of Central South University, Changsha, China

[2]Key Laboratory of Medical Information Research, Central South University, College of Hunan Province, Changsha, China

[3]School of Life Sciences, Central South University, Changsha, China

[4]School of Computer Science and Engineering, Central South University, Changsha, China

[5]Information and Network Center, Central South University, Changsha, China

Corresponding Author:
Zirui Xin, PhD
The Third Xiangya Hospital of Central South University
No.138 Tongzipo Road
Changsha, 410013
China
Phone: 86 0731 8861 8316
Email: xinzirui@csu.edu.cn

Abstract

Background: With the rapid development of online health communities, increasing numbers of patients and families are seeking health information on the internet.

Objective: This study aimed to discuss how to fully reveal the health information needs expressed by patients with hypertension in their questions in a web-based environment and how to use the internet to help patients with hypertension receive personalized health education.

Methods: This study randomly selected 1000 text records from the question data of patients with hypertension from 2008 to 2018 collected from Good Doctor Online and constructed a classification system through literature research and content analysis. This paper identified the background characteristics and questioning intention of each patient with hypertension based on the patient's question and used co-occurrence network analysis and the k-means clustering method to explore the features of the health information needs of patients with hypertension.

Results: The classification system for the health information needs of patients with hypertension included the following nine dimensions: drugs (355 names), symptoms and signs (395 names), tests and examinations (545 names), demographic data (526 kinds), diseases (80 names), risk factors (37 names), emotions (43 kinds), lifestyles (6 kinds), and questions (49 kinds). There were several characteristics of the explored web-based health information needs of patients with hypertension. First, more than 49% of patients described features, such as drugs, symptoms and signs, tests and examinations, demographic data, and diseases. Second, patients with hypertension were most concerned about treatment (778/1000, 77.80%), followed by diagnosis (323/1000, 32.30%). Third, 65.80% (658/1000) of patients asked physicians several questions at the same time. Moreover, 28.30% (283/1000) of patients were very concerned about how to adjust the medication, and they asked other treatment-related questions at the same time, including drug side effects, whether to take the drugs, how to treat the disease, etc. Furthermore, 17.60% (176/1000) of patients consulted physicians about the causes of clinical findings, including the relationship between the clinical findings and a disease, the treatment of a disease, and medications and examinations. Fourth, by k-means clustering, the questioning intentions of patients with hypertension were classified into the following seven categories: "how to adjust medication," "what to do," "how to treat," "phenomenon explanation," "test and examination," "disease diagnosis," and "disease prognosis."

Conclusions: In a web-based environment, the health information needs expressed by Chinese patients with hypertension to physicians are common and distinct, that is, patients with different background features ask relatively common questions to physicians. The classification system constructed in this study can provide guidance to health information service providers for the construction of web-based health resources, as well as guidance for patient education, which could help solve the problem of information asymmetry in communication between physicians and patients.

(J Med Internet Res 2020;22(5):e17349) doi:10.2196/17349

KEYWORDS

online health community; health information needs; patients with hypertension; physician-patient communication

Introduction

Background

Hypertension is currently the most common chronic disease and is a major risk factor for the morbidity and mortality of stroke and coronary heart disease in China [1]. According to statistics, the hypertension prevalence in Chinese residents aged 18 years or over is 27.9%, the hypertension awareness rate is 51.6%, and the hypertension control rate is 16.8% [2]. However, the hypertension control rate in the United States is close to 50% [3]. We urgently need to adopt measures to improve public awareness and health management of hypertension. An online health community is a health and medical online community of the *internet plus medical care* model, providing users with better knowledge services in the form of a community [4]. According to a research report, by the end of December 2018, the number of internet users in China had reached 820 million [5], and more than 190 million users had access to medical health information on the internet [6]. An online health community allows users to avoid long waits when seeking health information services, especially those with hypertension who need long-term self-health management at home. The health literacy of patients with hypertension in China is generally low. Owing to the small number of family physicians in the community and the low enthusiasm for signing contracts, the family physician policy has not been effectively promoted throughout the country [7]. After discharge, patients with hypertension cannot be guided or supervised by physicians to help manage the disease properly. When there are issues that may affect compliance, some patients will seek information or help on the internet. Wicks et al [8] showed that an online health community can help users benefit from managing their own disease symptoms, treating side effects, finding fellow patients, and asking for medication advice. We hope that information resource providers will effectively organize and recommend relevant information according to the information needs of each patient with hypertension. In the long term, this will help patients with hypertension reduce information asymmetry with physicians by personalized health education and improve their compliance. The research perspective of this study was to reflect the features of the information needs of patients with hypertension or their families in an internet-based environment according to the textual information disclosed by patients in an online health community when consulting with physicians.

Related Studies

At present, many studies have been carried out on the health information needs of internet users in China and other countries. The investigators mainly used questionnaires [9-11], in-depth interviews [12], and content analysis methods [13,14] to study the types of health information needs of specific populations. The study populations included elderly individuals [9], college students [15], pregnant and parturient women [16], and other populations, and the health problems involved were diabetes [17], hypertension [18], cancer [19], depression [20], and others. Related studies have shown that different populations have different web-based health information needs, which change with the course of the disease [10,12,21]. Patients with different cancers have high needs for information on the likelihood of cure, survival, treatment side effects, and risk [19].

In recent years, with the rapid development of natural language processing research and machine learning algorithms, methods based on topic recognition and text mining have been gradually applied to the analysis of web-based health information needs. Compared with the questionnaire survey method, the web text mining method can fully consider a patient's expression and idioms from a user's perspective. The obtained demand characteristics are more in line with the user's real information needs [22]. Chen used the k-means method to conduct clustering analysis on the text of three online health communities and found that different communities had different hot topics, as well as the same topics, such as patient experience, treatment, drugs, and body management [23]. Lv used the expectation maximization clustering method to perform topic analysis, member role analysis, and sentiment analysis on the MedHelp website and finally defined seven hot topics, including detailed personal introduction, emotional support, symptoms, examinations, complications, medication, and treatment [24]. Patrick et al [25] used a public dataset to predict the intent of the questioner and extracted patients' opinions, emotions, countermeasures, and social support topics from the data of patients with cancer and mental illness. Zhang [26] suggested that the background information of users, such as health status and age at the time of asking questions, should be considered in research on patient information needs. Owing to the professionalism and complexity of medical language and the issues of web users' descriptions, such as miswriting and colloquial words, it is difficult to manage user-generated content, and the machine learning algorithm model that it relies on needs to be trained and validated in an annotated corpus [27,28]. However, a Chinese corpus for patient consultation after current annotation is still lacking.

There are already some studies on the classification of health questions. However, these studies focused on information needs for health care professionals [26,28,29]. Guo et al [30,31]

constructed a Chinese health question classification corpus to study topics on the information needs of patients with hypertension, but these studies did not fully consider the background information of patients with hypertension, such as disease history, symptoms and signs, medication, gender, and age. As a result, they failed to reflect the features of the health information needs of more sophisticated patients with hypertension.

In summary, with the text information actually published in an online community as the research object, the obtained data can better reflect the real situation of the health information needs expressed by patients with hypertension during online consultations. However, nonmanual annotation methods, such as latent Dirichlet allocation (LDA) topic recognition, still have some deficiencies in Chinese medical text applications and cannot effectively reflect the semantic similarities of medical concepts. Although the LDA algorithm can identify the required topics from a large amount of text information, owing to the complexity and professionalism of the medical language, the feature keywords of the topic need to be manually summarized after generating the topic model. The final generated topic tags are medicine dependent, and expert discussions are designed to summarize the expressions of the most typical words [32]. Therefore, in this study, we used manual coding to analyze web text content.

Study Aim

In this study, on the basis of previous studies, we constructed a multidimensional classification system of the health information needs of patients with hypertension according to their consultation records in an online health community and used multidimensional features to reveal the features of the health information needs of patients with hypertension in a web-based environment.

Methods

Data Collection and Preprocessing

Data collection in this study was based on a search engine in the Chinese medical field [33] to filter nonreal, noneffective, and nonauthoritative medical information. Considering that the search engine can aggregate patient consultation data from different websites with disease keywords, it was convenient for us to obtain web-based consultation data for patients with hypertension. In addition, the search engine can mark the consulted physician, the physician's title, and the level of the medical institution. In this study, we searched for relevant records with *hypertension* as the keyword in the search engine and obtained 10,000 related records from 2008 to 2018. Most of the records provided by the search engine were from Good Doctor Online [34]. Good Doctor Online is one of China's leading web-based medical platforms. As of December 2019, it has information on 610,000 physicians from 9917 regular hospitals in China [34]. A survey of Good Doctor Online found that when consulting a physician, patients submit content according to a certain information description framework as follows: consultation title, disease, description of the condition, previous treatment status and effect, and how to get help. Considering that the resources under the same website have similar structures, in this study, we only included 8338 nonduplicate consultation records available on the website. A total of 1000 consultation records were randomly selected for further analysis, and the C# program (developed by YFY within the Key Laboratory of Medical Information Research, Central South University, College of Hunan Province, PR China) was used to extract information about the records of patient questions available on the website.

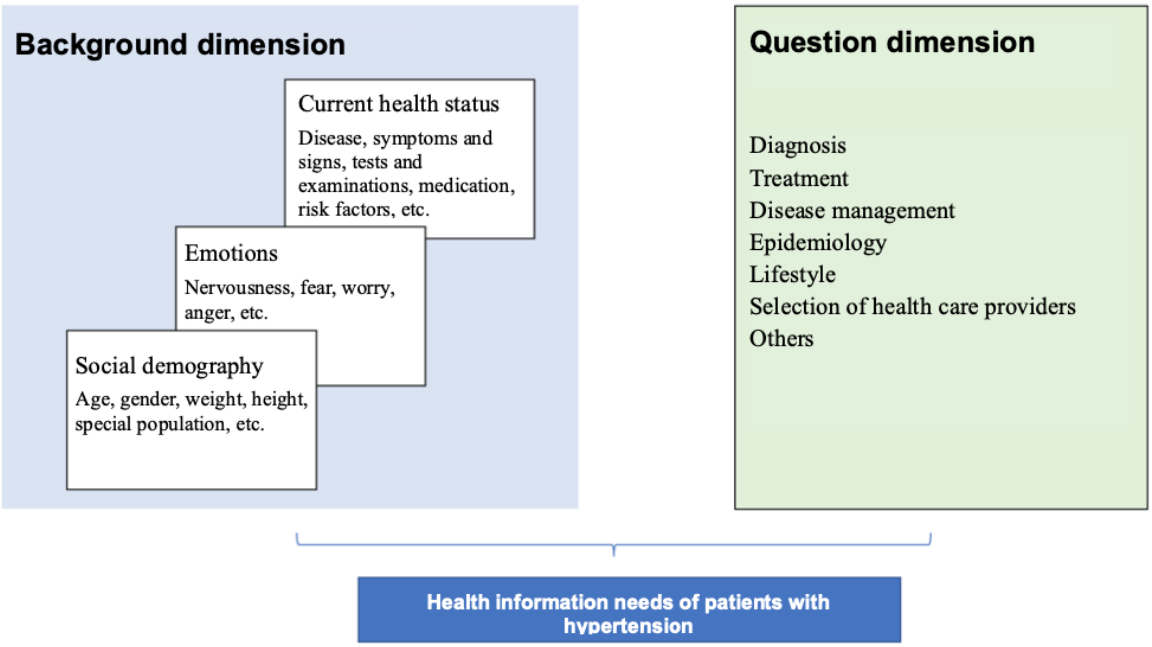
Multidimensional Classification System Construction

We initially summarized the knowledge framework of hypertension from the Guidelines for the Prevention and Treatment of Hypertension in China [35], the 8th Edition of Internal Medicine, and the Health Education Reader Book for Hypertensive Patients. Among them, the Guidelines for the Prevention and Treatment of Hypertension in China is the latest guidance document in China. With reference to the guideline development process of the World Health Organization and the Chinese Medical Association, the compilation team established a set of guidelines for the prevention and intervention, diagnosis and evaluation, classification and stratification, and treatment management of hypertension with Chinese characteristics according to the practice of hypertension prevention and treatment in China.

Based on the knowledge of hypertension, we initially constructed a system involving 12 categories, including epidemiology, etiology and pathogenesis, pathophysiology and pathology, clinical manifestations and complications, laboratory tests, diagnosis and differential diagnosis, prognosis, treatment, hypertension in special populations, risk factors, hazards, and management.

Thereafter, we took into consideration the above-mentioned general clinical question classification system [29], hierarchical model of public health information query scenario, and knowledge elements of hypertension prevention and treatment guidelines; read and summarized the consultation records (Figure 1); and initially constructed a multidimensional classification framework diagram for the health information needs of patients with hypertension. The classification framework mainly included the background dimension system (social demography, emotions, and current health status) and the question dimension system (diagnosis, treatment, disease management, epidemiology, lifestyle, selection of health care providers, and others).

Figure 1. Classification framework diagram for the health information needs of patients with hypertension.

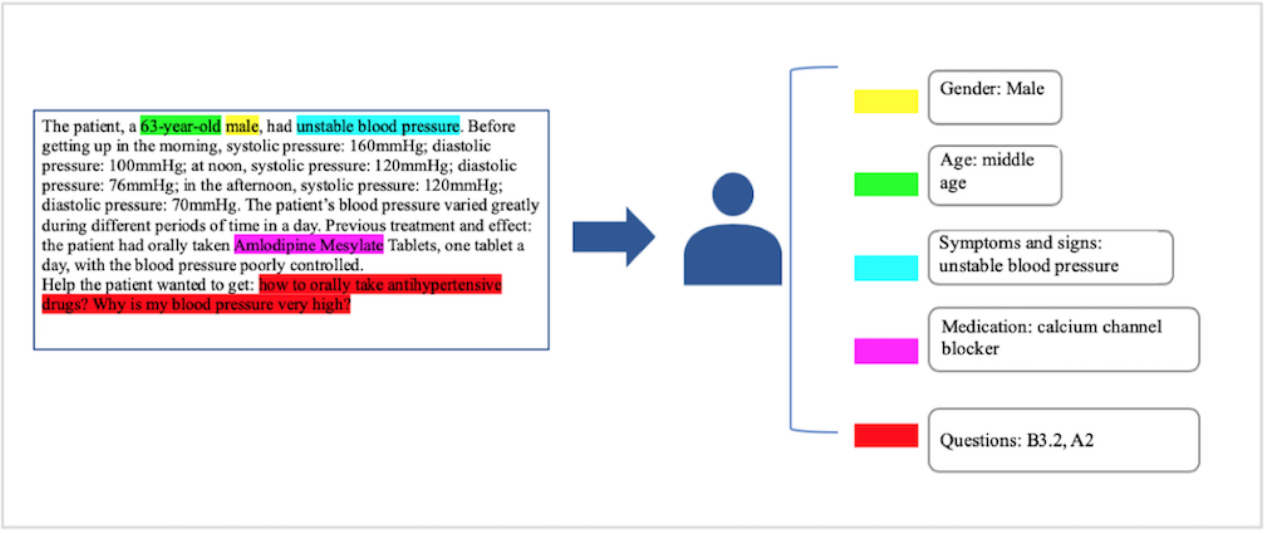


Annotation Process

Initially, two annotators with medical informatics backgrounds independently classified and annotated 100 records randomly selected from the 1000 consultation records mentioned according to the preliminary classification system and the classification example of each category. In the annotation process, each annotator proposed the experience of the annotation process of various entities, made the records, and finally completed the classification table and annotation descriptions for the health information needs of patients with hypertension after reaching a consensus through consultation. Thereafter, the two annotators randomly selected 200 records from the remaining records for repeat annotation, and they discussed inconsistent records (n=23) and reached a consensus after discussion. Finally, the two annotators completed the annotation of the remaining 700 questions and answers independently (ie, 350 questions and answers each). We manually aligned the entities of each type after the annotation was completed. Specifically, the original words were aligned with standard terms. An investigator with a background in medical informatics used this classification system for annotation. The classification system also included a series of annotation rules to improve the usability and accuracy of the classification system.

In particular, we coded each record and aligned the patient description terms with medical phrases under the joint guidance of two medical information professionals and one medical professional. We completed a structured representation of the original question records for patients with hypertension in the sample set. An example is shown in Figure 2.

Figure 2. Structured content coding for the original question record of a patient with hypertension. B3.2 denotes a question about taking medication and A2 denotes a question about symptoms and signs.



Co-Occurrence Analysis and Clustering Analysis

Co-occurrence analysis is one of the content analysis methods in bibliometrics. This method assumes that there is some inherent relationship between multiple keywords that appear in the same document. Co-occurrence frequency refers to the number of occurrences of a group of words in the same document. A larger co-occurrence frequency of two keywords indicates a closer relationship between the two words. Cluster analysis of these words based on co-occurrence frequency can reflect the association between multiple words [36]. Co-occurrence of keywords was considered to be the main approach for identifying research themes and knowledge structure by displaying the relationship between keywords [37]. This study used the co-occurrence analysis and clustering analysis methods to study the relationship between multiple questions from patients with hypertension on the internet. The representative k-means algorithm was selected in the cluster analysis.

This study used the integrated development environments PyCharm and RStudio under Windows, and Python3.6 and R language (developed by ZRX within the Key Laboratory of Medical Information Research, Central South University, College of Hunan Province, PR China) as programming languages for data analysis. In the co-occurrence analysis, the NumPy package was used to calculate the co-occurrence matrix of patient question codes, and the Circlize package was used to visually present the co-occurrence network between question codes. Besides, we used the Sklearn package to implement the k-means clustering algorithm in the cluster analysis.

Results

Feature Mining of the Background Dimension System

The background features of patients with hypertension can be divided into the following eight dimension types: disease, tests and examinations, symptoms and signs, drugs, risk factors, emotions, lifestyles, and demographic data. Among them, risk factors are summarized from the medical knowledge of hypertension, which intersects with other dimensions. The main contents include obesity, high salt and low potassium diet, genetics, diabetes, kidney disease, mental and psychological factors, smoking, drinking, pregnancy, hyperlipidemia, snoring, and others. The feature distribution of background dimensions in 1000 records is shown in Table 1. The drug dimension appeared most frequently in the sample set (814/1000, 81.40%), followed by symptoms and signs (795/1000, 79.50%), tests and examinations (563/1000, 56.30%), demographic data (527/1000, 52.70%), diseases (499/1000, 49.90%), risk factors (352/1000, 35.20%), emotions (119/1000, 11.90%), and lifestyles (114/1000, 11.40%).

Feature mining of the background dimension system can reflect the personalized features of the health information needs of every patient with hypertension on the internet. The personalized features cover 355 names of drugs, 395 names of symptoms and signs, 545 names of tests and examinations, 526 kinds of demographic data, 80 names of diseases, 37 names of risk factors, 43 kinds of emotions, 6 kinds of lifestyles, and 71 kinds of questions. In addition, for the demographic data, we built a system of rules. As shown in Table 2, based on the rules, we effectively mined a total of 385 patient gender records from the text (male: 178, female: 206), with gender unknown in other records. Moreover, we effectively mined a total of 352 age records, including seven underage patients, 177 young patients, 116 middle-aged patients, and 48 elderly patients. It can be seen that basic patient information disclosed on the internet is not complete, although it is important to mine patients' health information needs. To our knowledge, users currently ask physicians to provide sociodemographic information, such as age and gender, when consulting for disease-related issues on Good Doctor Online. However, this part is only visible to physicians and patients. Sociodemographic factors can be obtained by manually labeling the content according to the labeling rules from the public consultation data submitted by patients. However, the multidimensional classification system introduced in this study can make up for the incomplete or missing data of a single dimension through multidimensional mining.

Table 1. Frequency distribution of the background dimension system.

Dimension name	Keywords	Quantity	Records (N=1000), n (%)
Drugs	Calcium channel blocker, angiotensin II receptor antagonist, β blocker, angiotensin-converting enzyme inhibitor, traditional Chinese medicine, antihypertensive drugs, amlodipine besylate, Western medicine, compound antihypertensive drugs, diuretics, and aspirin enteric-coated tablets	355	814 (81.40%)
Symptoms and signs	Dizziness, headache, unstable blood pressure, chest discomfort, discomfort, giddiness, numbness, heart discomfort, pain, rapid heartbeat, movement disorder, palpitation, fatigue, weakness, and head swelling	395	795 (79.50%)
Tests and examinations	Physical examination, electrocardiogram examination, laboratory test, computed tomography examination, color Doppler ultrasound examination, cardiac color Doppler ultrasound examination, B-ultrasound examination, blood lipids, blood glucose, renal function, and routine blood test	545	563 (56.30%)
Demographic data	Gender, age, kinship, medical history, weight, height, fertility, occupation, family history, and others	526	527 (52.70%)
Diseases	Cerebrovascular disease (cerebral infarction, cerebral thrombosis, or cerebral hemorrhage), cardiovascular disease (coronary heart disease or heart disease), digestive system disease (fatty liver), rheumatic disease, other vascular disease, respiratory disease, endocrine disease, urinary tract systemic disease, and pregnancy-induced hypertension	80	499 (49.90%)
Risk factors	Dyslipidemia, diabetes, heredity, kidney disease, nervousness, fertility, drinking, obesity, stress, high blood glucose, smoking, emotional problems, anxiety, and menopause	37	352 (35.20%)
Emotions	Fear, doubt, worry, nervousness, vexation, irritability, anxiety, discomfort, exhaustion, moodiness, stress, bewilderment, depression, and emotional instability	43	119 (11.90%)
Lifestyles	Diet, exercise, sleep, weight loss, drinking, smoking, and others	6	114 (11.40%)

Table 2. Distribution of identified demographic data of the patients with hypertension.

Attribute	Quantity	Records (N=1000), n (%)
Gender		**385 (38.50%)**
Female	206	
Male	178	
Age		**352 (35.20%)**
Underage (<18 years)	7	
Young (18-44 years)	177	
Middle-aged (45-64 years)	116	
Elderly (over 65 years)	48	

Feature Mining of the Question Dimension System

The constructed question classification system for patients with hypertension included 49 items in seven categories, including diagnosis, treatment, disease management, epidemiology, lifestyle, selection of health care providers, and others. Owing to space limitations, only the distributions of the categories in the first and second classes of the classification dimension system are presented (Table 3).

Among 1000 web-based consultation records, 738 (73.80%) had patient questions that were related to treatment (B), mainly drug treatment (B3; quantity: 472), including drug selection, other drug-related questions, drug side effects, medication time, mode, and others. Among them, other drug-related questions were mainly about user consultation on whether medication is needed under the current situation. Regarding questions on medication time, there was a focus on long-term medication or drug discontinuation. Although it is widely known that patients with hypertension must take antihypertensive drugs for the rest of their life, they still asked this question repeatedly with anxiety and worry. Consultation about the side effects of drugs mainly involved drug safety, drug contraindications, and side effects.

Additionally, 32.30% (323/1000) of patients asked questions about the diagnosis of a disease (A), including seeking interpretation of clinical findings and recommendation of test items. Some patients also asked physicians questions about general diagnosis (A0) (eg, whether they contracted a disease). Moreover, 41.00% (410/1000) of patients consulted about not only medication (B1). Furthermore, 20.40% (204/1000) of patients asked questions about epidemiology (D1), such as the etiology of the disease and correlation between the disease and signs and symptoms, and their needs for information were not

limited to the diagnosis and treatment of the disease. Most medical workers paid little attention to answering these questions, and they mainly focused on the process of medical services. In addition, 4.80% (48/1000) of patients consulted about what they should pay attention to in daily life (E0) in order to maintain a good lifestyle, including diet (E4), weight loss (E1), exercise (E2), emotional management (E3), and many other aspects. Some patients realized that hypertension is a chronic disease and that a bad lifestyle could contribute to the progression of hypertension. They wanted to improve their health through a good lifestyle. Further, 10.00% (100/1000) of patients consulted about the selection of health care providers (F0), and they mainly focused on whether they should go to the hospital (F3) or choose hospitalization (F4) according to their physical conditions and other questions. Some patients asked questions about the impact of the disease on fertility (G1) and the reliability of their opinions. A total of 27 patients asked physicians about the reliability of opinions they saw on the internet or were told by others. Seven questions were classified into G0 (others-others) as they could not be classified into other categories.

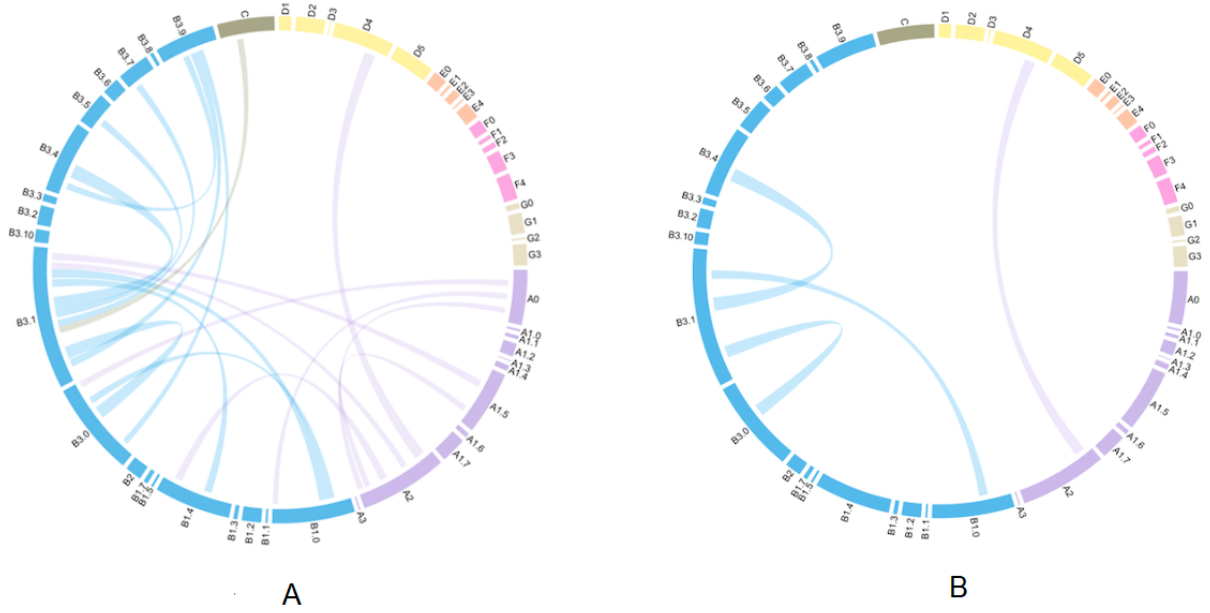
Table 3. Distribution of second-class codes of question features.

First category, second category	Quantity	Records (N=1000), n (%)
A: Diagnosis		**323 (32.30%)**
A0: Other diagnosis-related questions	97	
A1: Tests and examinations	113	
A2: Symptoms and signs	176	
A3: Diagnostic cost	2	
B: Treatment		**778 (77.80%)**
B1: Not only drug treatment	410	
B2: Nondrug treatment	29	
B3: Drug treatment	472	
C: Disease management	128	135 (13.50%)
D: Epidemiology		**204 (20.40%)**
D1: Etiology of a disease	27	
D2: Severity of a disease	41	
D3: Age at which a disease occurs	3	
D4: Others	74	
D5: Prognosis of a disease	75	
E: Lifestyle		**48 (4.80%)**
E0: Other lifestyles	18	
E1: Weight loss	4	
E2: Exercise	11	
E3: Emotion management	3	
E4: Diet	21	
F: Selection of health care providers		**103 (10.30%)**
F0: Other selection of health care provider-related questions	19	
F1: Department selection	7	
F2: Outpatient service	7	
F3: Physician selection	37	
F4: Hospital selection	40	
G: Others		**66 (6.60%)**
G0: Others	7	
G1: Fertility	31	
G2: Medical insurance	2	
G3: Opinion judgment	27	

Feature Mining of Co-Occurrence in the Question Dimension System

Since each patient's online consultation contains multiple questions and involves various information needs, co-occurrence network analysis based on the question type was conducted according to the question features. A total of 658 patients with hypertension asked physicians several questions from one online consulting record. Treatment (B) and diagnosis (A) were the most common issues among patients (Figure 3). Because patients with hypertension need to take drugs for a long time, 283 patients who consulted physicians on the internet were very concerned about drug selection (B3.1). In addition, they asked other treatment-related questions (B), including side effects (B3.4), whether to take the drug (B3.0), how to treat the disease (B1.0), etc. Moreover, among 176 patients who consulted physicians on the internet about the causes of clinical findings (A2), 39 consulted physicians about the relationship between clinical findings and the disease (D4).

Figure 3. Co-occurrence graph for question type. (A) Co-occurrence frequency ≥20%; (B) Co-occurrence frequency ≥30%.



Feature Mining for Patients With Hypertension According to Background and Question Dimensions

We performed k-means clustering on all classification features of the consultation. After many tests, we selected question type as the feature variable, and the elbow method determined that the clustering model was the best when k was 7. Finally, we obtained a schematic diagram of seven types of questions for patients with hypertension. The medical concepts expressed by the main features of each type were clearly defined (Table 4). The seven categories of the population contained background details, such as test and examination items, dizziness symptoms, and calcium channel blocker use.

The first category of the population was "how to adjust medication," which included patients who mainly needed information about drug selection. The background details of their questions mainly involved information on the names of drugs, including calcium channel blockers, angiotensin II receptor antagonists, angiotensin converting enzyme inhibitors, and β-receptor blockers. The second category was "what to do," which included patients who mainly asked physicians to provide information on disease management and prevention. There were more female patients in this group. The third category was "how to treat," which included patients who were mainly seeking information from physicians on disease treatment. Their background descriptions focused on the use of calcium channel blockers and dizziness symptoms. The fourth category was "phenomenon explanation," which included patients who consulted physicians about the causes of clinical findings, the correlation between clinical findings and a certain disease, and the treatment plan of the disease. Their background descriptions mainly included tests and examinations, and dizziness and headache. Most of them were female patients who used calcium channel blockers and angiotensin II receptor antagonists. The fifth category was "test and examination," which included patients who mainly consulted physicians about tests and examinations. Their background data mainly included the description of tests and examinations, and dizziness. Most of them were young male patients who used calcium channel blockers. The sixth category was "disease diagnosis," which included patients who mainly consulted physicians about whether they had a certain disease, whether they needed drugs, the causes of symptoms and signs, and their treatment. Their background data mainly included information of tests and examinations, and dizziness. Most of them were young patients. The seventh category was "disease prognosis," which included patients who mainly asked physicians questions such as "Can the disease be cured?" and "How to treat it?" Their background data mainly included tests and examinations, dizziness, brain disease, and calcium channel blocker use.

Table 4. Classification of the health information needs of patients with hypertension.

Category	Main question, category statistics (N=1000)	Examples of typical questions	Description of main background features (N=1000)	Quantity
1	B3.1 (n=185); B3.0 (71); B3.4 (61)	How to adjust medication or change drugs? Do I need medication? Does the medicine have side effects?	Calcium channel blockers (n=124), angiotensin II receptor antagonists (74), angiotensin-converting enzyme inhibitors (61), dizziness (59), tests and examinations (56), and β-receptor blockers (54)	262
2	C (69)	What to do and how to prevent it?	Tests and examinations (64), female (53), dizziness (47), and calcium channel blockers (45)	205
3	B1.0 (110); B1.4 (68)	How to treat it? How to treat a disease?	Calcium channel blockers (48) and dizziness (46)	184
4	A2 (97); D4 (24); B1.4 (20)	Why do I have such a symptom? What is the relationship between a disease and a symptom? How to treat a disease?	Tests and examinations (36), dizziness (31), female (28), calcium channel blockers (28), headache (24), and angiotensin II receptor antagonists (23)	97
5	A1.5 (53); A1.7 (31)	What examinations are needed? What does the result of the examination mean?	Tests and examinations (66), male (26), dizziness (26), youth (19), and calcium channel blockers (19)	94
6	A0 (85); B3.0 (19); A2 (18); B1.0 (18)	Have I got a disease? Do I need medication? Why do I have such a symptom? How to treat it?	Tests and examinations (32), dizziness (21), and youth (19)	85
7	D5 (73); B1.0 (16); B1.4 (15)	Can this disease be cured? How to treat it? How to treat this disease?	Tests and examinations (29), dizziness (19), brain disease (19), and calcium channel blockers (17)	73

Discussion

Principal Findings

We found some common and distinct features of information needs from the consultation records of patients with hypertension in an online health community. In this study, we constructed a multidimensional *background plus question* classification system for mining various features of health information needs and effectively revealed what Chinese patients with hypertension want to communicate with physicians online. According to our knowledge, no researchers have studied the classification of health information needs in background and question dimensions among patients with hypertension. Our constructed classification system underlines the personal features of the information needs of patients with hypertension, which could help online information service providers and relevant researchers personalize their services. Related online health information resources can be identified according to the classification system of hypertension information needs in this study so as to make online health information resource organization more refined and meet the needs of users.

We found that drug information is quite important to Chinese patients with hypertension on the internet. Additionally, 81.40% (814/1000) of them described their previous medication situation when consulting physicians on the internet, and there are many types of drugs involved. Moreover, 73.80% (738/1000) of them asked their physicians how to treat the disease. This is similar to the findings of Guo et al [30,31,38]. Chinese patients with hypertension mainly asked physicians about drug treatment, including drug selection, drug side effects, medication time, and mode. The results of Zhong et al [39] also showed that the medication literacy of discharged patients with hypertension was poor. More than 70% of patients had no substantial knowledge of the effects and side effects of the medications they were taking. It is worth noting that 16.10% (161/1000) of the questioners with hypertension were concerned about deeper information, including the prognosis, etiology, and progression of the disease.

Most background information described by patients with hypertension when they consulted a physician on the internet involved drug information, followed by symptoms and signs, which accounted for about 80% of patient records. More than 49% of patients mentioned details, such as tests and examinations, diseases, and sociodemographic data. To our knowledge, these values have only been revealed in our study. In this study, we found that patients with hypertension were most concerned about treatment (738/1000, 73.80%) and diagnosis (313/1000, 31.30%). This is similar to the findings of Guo et al [38]. It is worth noting that 16.10% (161/1000) of patients with hypertension were concerned about deeper information, such as the prognosis, etiology, and progression of the disease. In a web-based consultation, Chinese patients with hypertension usually ask physicians many questions without significant priorities, but in an offline consultation, physicians often fail to answer some questions patients are

concerned about owing to the lack of time and the long line of waiting patients.

In this study, we also found common associations when patients asked questions. This is different from the findings of Guo et al [30], and it has not been reported in previous studies. Clustering results based on question features can be explained by medical knowledge. Patients with hypertension can be classified into seven categories, including "how to adjust medication," "what to do," "how to treat," "phenomenon explanation," "test and examination," "disease diagnosis," and "disease prognosis." When new patients with hypertension ask questions, health information service providers can predict which type of population the patients belong to using the clustering model and can recommend other relevant information that the patients may need while recommending the question information. However, owing to the complexity and uniqueness of questioners' background information, we tried to combine the background and question features to conduct clustering research and association rule mining research; however, we did not obtain an interpretable clustering model. This also reflects the personalized differences in the real health information needs of patients with hypertension.

Some patients with hypertension in China still lack an accurate understanding of the risk factors of hypertension and need to acquire more effective knowledge. The risk factors for hypertension were involved in 35.00% (350/1000) of consultations in this study. The risk factors appearing in the research results were presented by the patients when describing the problem, and they may not necessarily recognize the factors as risk factors. Those risk factors were also not question intentions at that time. We hope that when patients consult physicians on the internet, future internet tools will be able to highlight the risk factors of patients with hypertension in a timely manner, as well as promptly remind and push health intervention knowledge of corresponding diseases and their risk factors to patients with hypertension. This is conducive to the provision of targeted information services by information resource providers and to meet the personalized goals of patient education. For example, it can promptly and continuously alert obese patients with hypertension to take measures to control their weight.

The patients with hypertension included underage, young, middle-aged, and elderly people, and thus, hypertension in China can occur in people of all ages. In future health information services, the information that is more in line with the actual health situation and that can improve the health status could be screened from dimensions, such as population and risk factors. In addition, some women were worried about the effects of drugs on pregnancy when they asked about medication. At present, the second-child policy has been widely implemented in China. Medical service personnel need to pay special attention to pregnant women, such as women with hypertension, help them give birth to healthy babies under the support of the existing medical resources, strengthen their accurate understanding of hypertension from the perspective of health information services, and help them relieve their anxiety.

In our study, we found that manual consultation analysis of patient consultation text in Chinese is still needed. Among patient consultation records, medical terms, such as drugs, symptoms and signs, tests and examinations, and diseases, are often not standardized in the patient information description. Owing to typographical errors, omissions, ambiguous words, etc, manual alignment or correction is still needed. This was also noted in the article by Guo et al [30]. The glossary of terms accumulated in this study on summarizing the wording habits of patients in consultations will improve the accuracy of machine learning in feature recognition so as to obtain more accurate predictions of needs categories.

Through this study, we hope to strengthen patient participation with the help of internet technology for improving the personalized characteristics of existing patient education. Patient engagement in their care plays an important role in improving health outcomes [40]. With patient participation, patient education can provide detailed information about drugs, symptoms and signs, tests and examinations, demographic data, diseases, risk factors, emotions, lifestyles, and other information identified in the consultation records for each patient in real time. The cluster analysis method extracts the corresponding population characteristics of patients and further provides the information that the patients need or may need. The process of providing health information to patients through a platform has actually begun to provide health education to the patients, that is, when patients are puzzled about health problems and submit questions on the internet, they can respond to health education services with the help of network technology. Personal questioning data have accumulated on web-based platforms, which have formed the growing consulting big data of users in online health communities. With the help of these big data and the cluster analysis method, it is possible to identify the subdivided hypertension group and obtain the generally needed health knowledge to help the construction and continuous update of patient health education content. We hope that in the future, internet tools will become individual physician assistants for patients under the supervision of professionals. For the information needs of patients, health education can be effectively carried out in real time.

In this study, we used a multidimension approach to identify the features of the health information needs of patients with hypertension in a web-based environment when consulting a physician. Thereafter, according to the classification of these needs, we effectively organized and constructed web-based health information resources and accurately matched the web-based health information resources with the health information needs of each questioner through identification. The unmet information needs will be the focus of future web-based health information resources. In future studies, we will determine how to assess the degree of matching between health information needs and web-based health information resources. In the future, we will also determine how to apply this classification system to clinical consultations in order to reduce information asymmetry between physicians and patients and to help patients and physicians manage diseases together.

Limitations

The classification system may have limitations owing to the insufficient sample size for analysis. In the future, we should continue to expand the sample size and carry out a larger classification of health information needs. The sample data were from a single website, and all information was from the web-based consultation records available on the website. Thus, the target group of the website may affect the results of the construction of the classification system. In the future, we will collect data from several health websites and conduct comparative studies. The information needs of patients with hypertension in this study were the information needs expressed by patients. However, this study did not fully address the feature recognition of patients' unexpressed information needs. Our next goal is to determine how to improve compliance in patients with hypertension.

Conclusions

The classification system of hypertension information needs involving the *background plus question* approach constructed in this study reflects the personalized features of the health information needs expressed by patients with hypertension to physicians in a web-based environment. We believe that multidimensional health information needs should have more refined identifiers, which will enable patients with hypertension to fully express their health information needs during a web-based consultation, thus allowing relevant information service providers of online health communities to organize information resources related to hypertension according to these identifiers. They can use the classification system of hypertension information needs constructed in this study as a classification basis to organize web-based health information resources and provide more accurate information push services to patients when they consult physicians over the internet or wait for physicians to reply in order to reduce information asymmetry in communication between physicians and patients. On the other hand, the Chinese corpus of patients with hypertension based on artificial coding in this study will lay the foundation for future research on the features of automatic identification of health information.

Acknowledgments

We thank the State Key Program of National Social Science of China (grant no. 17AZD037) for assistance in this study. This work was partially supported by the National Natural Science Foundation of China (grant no. 61977062) and the Social Science Foundation in Hunan Province (grant no. 16YBA366). We also appreciate the support of all participants.

Authors' Contributions

AJL conducted the project and designed this research. ZRX drafted the manuscript and programmed with Python and R language. YFY implemented the data collection, preprocessing, and alignment modules. ZQZ, XQP, WZX, WOY, CH, and HYH identified the classification framework. YFY and FL were involved in annotation work. ZQZ and XQP provided their medical expertise. TXW and YC helped analyze the data. All authors reviewed and approved the manuscript for publication.

Conflicts of Interest

None declared.

References

1. Guo QH, Zhang YQ, Wang JG. Asian management of hypertension: Current status, home blood pressure, and specific concerns in China. J Clin Hypertens (Greenwich) 2020 Mar;22(3):475-478 [FREE Full text] [doi: 10.1111/jch.13687] [Medline: 31622005]
2. Wang Z, Chen Z, Zhang L, Wang X, Hao G, Zhang Z, China Hypertension Survey Investigators. Status of Hypertension in China: Results From the China Hypertension Survey, 2012-2015. Circulation 2018 May 29;137(22):2344-2356. [doi: 10.1161/CIRCULATIONAHA.117.032380] [Medline: 29449338]
3. Whelton PK, Carey RM, Aronow WS, Casey DE, Collins KJ, Dennison Himmelfarb C, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Circulation 2018 Oct 23;138(17):e484-e594. [doi: 10.1161/CIR.0000000000000596] [Medline: 30354654]
4. Liu W, Fan X, Ji R, Jiang Y. Perceived Community Support, Users' Interactions, and Value Co-Creation in Online Health Community: The Moderating Effect of Social Exclusion. Int J Environ Res Public Health 2019 Dec 27;17(1):204 [FREE Full text] [doi: 10.3390/ijerph17010204] [Medline: 31892188]
5. China Internet Network Information Center. The 43rd China Statistical Report on Internet Development URL: http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/201701/P020170123364672657408.pdf [accessed 2019-09-20]
6. China Internet Network Information Center. The 39rd China Statistical Report on Internet Development URL: http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/201701/P020170123364672657408.pdf [accessed 2019-09-20]

7. Yuan S, Wang F, Li X, Jia M, Tian M. Facilitators and barriers to implement the family doctor contracting services in China: findings from a qualitative study. BMJ Open 2019 Oct 08;9(10):e032444-e032299 [FREE Full text] [doi: 10.1136/bmjopen-2019-032444] [Medline: 31597653]
8. Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, et al. Sharing health data for better outcomes on PatientsLikeMe. J Med Internet Res 2010;12(2):e19 [FREE Full text] [doi: 10.2196/jmir.1549] [Medline: 20542858]
9. 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]
10. Zhang J. Supporting information needs of transitional phases in diabetes management through online health communitie. 2017 Presented at: Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing; February 2017; Oregon, USA p. 107-111. [doi: 10.1145/3022198.3024942]
11. Wellburn S, van Schaik P, Bettany-Saltikov J. The Information Needs of Adolescent Idiopathic Scoliosis Patients and Their Parents in the UK: An Online Survey. Healthcare (Basel) 2019 Jun 20;7(2):78 [FREE Full text] [doi: 10.3390/healthcare7020078] [Medline: 31226769]
12. Rozmovits L, Ziebland S. What do patients with prostate or breast cancer want from an Internet site? A qualitative study of information needs. Patient Education and Counseling 2004 Apr;53(1):57-64. [doi: 10.1016/s0738-3991(03)00116-2]
13. Cheng FW, So SW, Fung BW, Hung WH, Lee VW. Online drug information platform for the public in Hong Kong-Review of local drug information use and needs. Int J Med Inform 2018 Jun;114:27-34. [doi: 10.1016/j.ijmedinf.2018.03.006] [Medline: 29673600]
14. Suziedelyte A. How does searching for health information on the Internet affect individuals' demand for health care services? Soc Sci Med 2012 Nov;75(10):1828-1835. [doi: 10.1016/j.socscimed.2012.07.022] [Medline: 22884947]
15. Britt RK, Collins WB, Wilson K, Linnemeier G, Englebert AM. eHealth Literacy and Health Behaviors Affecting Modern College Students: A Pilot Study of Issues Identified by the American College Health Association. J Med Internet Res 2017 Dec 19;19(12):e392 [FREE Full text] [doi: 10.2196/jmir.3100] [Medline: 29258979]
16. Oscarsson MG, Medin E, Holmström I, Lendahls L. Using the Internet as source of information during pregnancy - a descriptive cross-sectional study among fathers-to-be in Sweden. Midwifery 2018 Jul;62:146-150. [doi: 10.1016/j.midw.2018.04.008] [Medline: 29684793]
17. Biernatzki L, Kuske S, Genz J, Ritschel M, Stephan A, Bächle C, et al. Information needs in people with diabetes mellitus: a systematic review. Syst Rev 2018 Feb 14;7(1):27 [FREE Full text] [doi: 10.1186/s13643-018-0690-0] [Medline: 29444711]
18. Lynch HM, Green A, Clarke Nanyonga R, Gadikota-Klumpers DD, Squires A, Schwartz JI, et al. Exploring patient experiences with and attitudes towards hypertension at a private hospital in Uganda: a qualitative study. Int J Equity Health 2019 Dec 30;18(1):206 [FREE Full text] [doi: 10.1186/s12939-019-1109-9] [Medline: 31888767]
19. Alba-Ruiz R, Bermúdez-Tamayo C, Pernett JJ, Garcia-Gutierrez JF, Cózar-Olmo JM, Valero-Aguilera B. Adapting the content of cancer web sites to the information needs of patients: reliability and readability. Telemed J E Health 2013 Dec;19(12):956-966 [FREE Full text] [doi: 10.1089/tmj.2013.0050] [Medline: 24073899]
20. Barney LJ, Griffiths KM, Banfield MA. Explicit and implicit information needs of people with depression: a qualitative investigation of problems reported on an online depression support forum. BMC Psychiatry 2011 May 18;11(1). [doi: 10.1186/1471-244x-11-88]
21. Qiu JJ, Li P. Information of patients with breast cancer in different stages and their care needs. Nursing Research of China 2016;30(15):1823-1827. [doi: 10.3969/j.issn.1009-6493.2016.15.011]
22. Hao H, Zhang K. The Voice of Chinese Health Consumers: A Text Mining Approach to Web-Based Physician Reviews. J Med Internet Res 2016 May 10;18(5):e108 [FREE Full text] [doi: 10.2196/jmir.4430] [Medline: 27165558]
23. Chen AT. Exploring online support spaces: using cluster analysis to examine breast cancer, diabetes and fibromyalgia support groups. Patient Educ Couns 2012 May;87(2):250-257. [doi: 10.1016/j.pec.2011.08.017] [Medline: 21930359]
24. Lv Y, Zhang P, Liu J, Li J, Deng S. Health-related hot topic detection in online communities using text clustering. PLoS One 2013;8(2):e56221 [FREE Full text] [doi: 10.1371/journal.pone.0056221] [Medline: 23457530]
25. Patrick J, Li M. An ontology for clinical questions about the contents of patient notes. J Biomed Inform 2012 Apr;45(2):292-306 [FREE Full text] [doi: 10.1016/j.jbi.2011.11.008] [Medline: 22142949]
26. Zhang Y. Toward a layered model of context for health information searching: An analysis of consumer-generated questions. J Am Soc Inf Sci Tec 2013 Apr 05;64(6):1158-1172. [doi: 10.1002/asi.22821]
27. Kilicoglu H, Ben Abacha A, Mrabet Y, Shooshan SE, Rodriguez L, Masterton K, et al. Semantic annotation of consumer health questions. BMC Bioinformatics 2018 Feb 06;19(1):34 [FREE Full text] [doi: 10.1186/s12859-018-2045-1] [Medline: 29409442]
28. Roberts K, Demner-Fushman D. Interactive use of online health resources: a comparison of consumer and professional questions. J Am Med Inform Assoc 2016 Jul;23(4):802-811. [doi: 10.1093/jamia/ocw024] [Medline: 27147494]
29. Ely JW, Osheroff JA, Gorman PN, Ebell MH, Chambliss ML, Pifer EA, et al. A taxonomy of generic clinical questions: classification study. BMJ 2000 Aug 12;321(7258):429-432 [FREE Full text] [doi: 10.1136/bmj.321.7258.429] [Medline: 10938054]

30. Guo H, Na X, Hou L, Li J. Classifying Chinese Questions Related to Health Care Posted by Consumers Via the Internet. J Med Internet Res 2017 Jun 20;19(6):e220 [FREE Full text] [doi: 10.2196/jmir.7156] [Medline: 28634156]
31. Guo H, Na X, Li J. Qcorp: an annotated classification corpus of Chinese health questions. BMC Med Inform Decis Mak 2018 Mar 22;18(Suppl 1):16 [FREE Full text] [doi: 10.1186/s12911-018-0593-y] [Medline: 29589562]
32. Hao H, Zhang K. The Voice of Chinese Health Consumers: A Text Mining Approach to Web-Based Physician Reviews. J Med Internet Res 2016 May 10;18(5):e108 [FREE Full text] [doi: 10.2196/jmir.4430] [Medline: 27165558]
33. SOGOU. URL: https://mingyi.sogou.com/ [accessed 2020-04-30]
34. Good Doctor Online. Good Doctor Online Introduction URL: https://www.haodf.com/info/aboutus.php [accessed 2020-02-01]
35. Joint Committee for Guideline Revision. 2018 Chinese Guidelines for Prevention and Treatment of Hypertension-A report of the Revision Committee of Chinese Guidelines for Prevention and Treatment of Hypertension. J Geriatr Cardiol 2019 Mar;16(3):182-241 [FREE Full text] [doi: 10.11909/j.issn.1671-5411.2019.03.014] [Medline: 31080465]
36. Li F, Li M, Guan P, Ma S, Cui L. Mapping publication trends and identifying hot spots of research on Internet health information seeking behavior: a quantitative and co-word biclustering analysis. J Med Internet Res 2015 Mar 25;17(3):e81 [FREE Full text] [doi: 10.2196/jmir.3326] [Medline: 25830358]
37. Yang A, Lv Q, Chen F, Wang D, Liu Y, Shi W. Identification of Recent Trends in Research on Vitamin D: A Quantitative and Co-Word Analysis. Med Sci Monit 2019 Jan 22;25:643-655 [FREE Full text] [doi: 10.12659/MSM.913026] [Medline: 30668558]
38. Guo H, Li J, Dai T. Consumer Health Information Needs and Question Classification: Analysis of Hypertension Related Questions Asked by Consumers on a Chinese Health Website. Stud Health Technol Inform 2015;216:810-814. [Medline: 26262164]
39. Zhong Z, Ma G, Zheng F, Duan Y, Ding S, Luo A. Medication Literacy in a Cohort of Chinese Patients Discharged With Essential Hypertension. Front Public Health 2019 Jan 9;7:385 [FREE Full text] [doi: 10.3389/fpubh.2019.00385] [Medline: 31998676]
40. Shetty KR, Wang RY, Shetty A, Levi J, Aaronson NL. Quality of Patient Education Sections on Otitis Media Across Different Website Platforms. Ann Otol Rhinol Laryngol 2020 Jun;129(6):591-598. [doi: 10.1177/0003489420902183] [Medline: 31975608]

Abbreviations

LDA: latent Dirichlet allocation

Edited by G Eysenbach; submitted 09.12.19; peer-reviewed by W Fuzhi, J Chudek; comments to author 10.01.20; revised version received 16.02.20; accepted 29.03.20; published 29.05.20.

Please cite as:
Luo A, Xin Z, Yuan Y, Wen T, Xie W, Zhong Z, Peng X, Ouyang W, Hu C, Liu F, Chen Y, He H
Multidimensional Feature Classification of the Health Information Needs of Patients With Hypertension in an Online Health Community Through Analysis of 1000 Patient Question Records: Observational Study
J Med Internet Res 2020;22(5):e17349
URL: http://www.jmir.org/2020/5/e17349/
doi:10.2196/17349
PMID:32469318

Original Paper

Distinguishing Obstructive Versus Central Apneas in Infrared Video of Sleep Using Deep Learning: Validation Study

Sina Akbarian[1,2,3], MASc; Nasim Montazeri Ghahjaverestan[1,2], PhD; Azadeh Yadollahi[1,2], PhD; Babak Taati[1,2,3,4], PhD

[1]Kite Research Institute, Toronto Rehabilitation Institute, University Health Network, Toronto, ON, Canada

[2]Institute of Biomaterials & Biomedical Engineering, University of Toronto, Toronto, ON, Canada

[3]Vector Institute, Toronto, ON, Canada

[4]Department of Computer Science, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Babak Taati, PhD
Kite Research Institute, Toronto Rehabilitation Institute
University Health Network
550 University Ave
Toronto, ON
Canada
Phone: 1 4165973422 ext 7972
Email: babak.taati@uhn.ca

Abstract

Background: Sleep apnea is a respiratory disorder characterized by an intermittent reduction (hypopnea) or cessation (apnea) of breathing during sleep. Depending on the presence of a breathing effort, sleep apnea is divided into obstructive sleep apnea (OSA) and central sleep apnea (CSA) based on the different pathologies involved. If the majority of apneas in a person are obstructive, they will be diagnosed as OSA or otherwise as CSA. In addition, as it is challenging and highly controversial to divide hypopneas into central or obstructive, the decision about sleep apnea type (OSA vs CSA) is made based on apneas only. Choosing the appropriate treatment relies on distinguishing between obstructive apnea (OA) and central apnea (CA).

Objective: The objective of this study was to develop a noncontact method to distinguish between OAs and CAs.

Methods: Five different computer vision-based algorithms were used to process infrared (IR) video data to track and analyze body movements to differentiate different types of apnea (OA vs CA). In the first two methods, supervised classifiers were trained to process optical flow information. In the remaining three methods, a convolutional neural network (CNN) was designed to extract distinctive features from optical flow and to distinguish OA from CA.

Results: Overnight sleeping data of 42 participants (mean age 53, SD 15 years; mean BMI 30, SD 7 kg/m^2; 27 men and 15 women; mean number of OA 16, SD 30; mean number of CA 3, SD 7; mean apnea-hypopnea index 27, SD 31 events/hour; mean sleep duration 5 hours, SD 1 hour) were collected for this study. The test and train data were recorded in two separate laboratory rooms. The best-performing model (3D-CNN) obtained 95% accuracy and an F_1 score of 89% in differentiating OA vs CA.

Conclusions: In this study, the first vision-based method was developed that differentiates apnea types (OA vs CA). The developed algorithm tracks and analyses chest and abdominal movements captured via an IR video camera. Unlike previously developed approaches, this method does not require any attachment to a user that could potentially alter the sleeping condition.

(J Med Internet Res 2020;22(5):e17252) doi:10.2196/17252

KEYWORDS

noncontact monitoring; sleep apnea; motion analysis; computer vision; obstructive apnea; central apnea; machine learning; deep learning

Introduction

Background

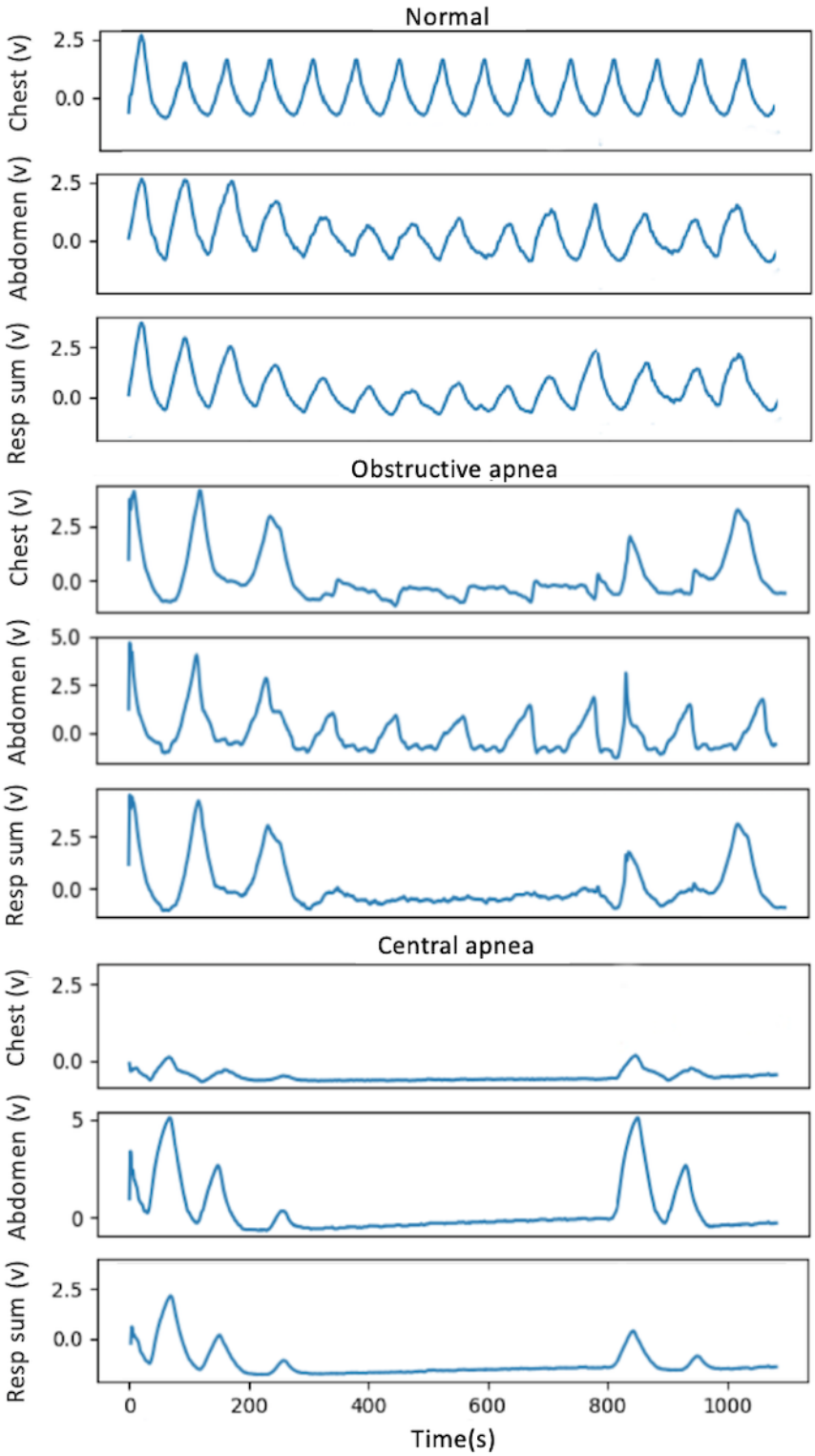
Sleep apnea is a chronic respiratory disorder, caused by intermittent reduction (hypopnea) or cessation (apnea) of respiratory airflow during sleep. About 10% of the population have this disorder [1], and it increases the risk of heart disease by 3-fold, stroke by 4-fold, and car accidents by 7-fold [2-5]. The severity of sleep apnea is commonly measured via the apnea-hypopnea index (AHI), which shows the number of apneas and hypopneas per hour of sleep.

Depending on the presence of breathing effort, sleep apneas can be divided into obstructive sleep apnea (OSA) or central sleep apnea (CSA) by measuring the thoracoabdominal movement and its contributions to the total respiratory volume [6]. The majority of sleep apneas are obstructive [7], which is caused by the full collapse of the pharyngeal airway that blocks the flow of air into the lungs [8]. The rest of sleep apneas are central, which happen due to a reduction in the respiratory drive from the central nervous system [9].

During normal breathing, the chest and abdomen move in phase due to interaction between the diaphragm and parasternal intercostals during inhalation [10]. During obstructive apnea (OA), airway obstruction leads to an out-of-phase motion of the rib cage and the abdomen, causing a reduction in the sum of chest and abdomen's movement [6]. On the other hand, during central apnea (CA), there is no movement in the rib cage or the abdomen due to lack of brain signal for muscle contraction. Figure 1 compares the movements in the chest and the abdomen and the sum of them during OA, CA, and normal breathing. If the majority of events in a person are obstructive, they will be diagnosed as OSA and otherwise as CSA. In addition, as it is challenging and highly controversial to divide hypopneas into central or obstructive, the decision about sleep apnea type (OSA vs CSA) is made based on apneas only.

The treatment of sleep apnea decreases patients' health-related costs by 25% [11]. A highly effective treatment for OSA is continuous positive airway pressure (CPAP) therapy that involves applying positive pressure of air to keep the airway open during sleep. However, CPAP therapy leads to increase mortality in patients with CSA [12]. Therefore, a crucial step for proper treatment of sleep apnea is to differentiate OSA from CSA [13].

Figure 1. Difference between the movements of the chest and abdomen and the sum of the two movements during obstructive apnea (OA), central apnea (CA), and normal breathing. During normal breathing, chest and abdomen movements are in phase. During OA, breathing effort and airway blockage result in the out-of-phase movement of the chest and abdomen, and the sum of the two movement signals (respiratory sum) is close to zero. During CA, there is no movement in the chest or the abdomen. CA: central apnea; OA: obstructive apnea; Resp sum: respiratory sum.



Problem Statement

The gold standard for differentiating OA from CA is the evaluation of swings in esophageal pressure measured via an esophageal catheter [14]. Measurement of esophageal pressure is invasive, uncomfortable, and could disrupt sleep. Therefore, esophageal pressure measurements are only used for physiological research purposes [15,16]. The current clinical approach to distinguish OA from CA based on the different patterns of movement in the chest and abdomen is respiratory inductance plethysmography (RIP). It measures thoracoabdominal movements from two transducer bands over the chest and the abdomen. RIP is a part of polysomnography (PSG), which consists of connecting more than 20 sensors to

patients. PSG is inconvenient during sleep, expensive (>US $400 in Canada), has long waiting time (4-36 months in Canada [17]), and requires a trained technician to manually analyze the recorded signals, including RIP.

A promising approach to distinguish different types of sleep apnea is using computer vision systems. The goal of computer vision is to extract information from images or videos. In this study, a noncontact approach was proposed to identify the types of apneas (OA vs CA) using computer vision techniques. In this approach, the upper body respiratory movements were captured in infrared (IR, *night vision*) video frames and were processed via computer vision algorithms.

Computer vision systems have been previously developed for various applications in sleep monitoring, for example, to monitor the breathing rate and pulse [18-20], to estimate the AHI [21], to screen sleep quality and irregularities [22,23], and to monitor sleep positions [24]. But the use of computer vision to distinguish different types of apnea (OA vs CA) was unexplored and was the main novel contribution of this study.

Methods

Data Collection

Adults aged 18 years or above who were referred to the sleep laboratory of the Toronto Rehabilitations Institute-University Health Network for sleep screening were recruited for this study. The University Health Network Research Ethics Board approved this study (Research Ethics Board approval number 13-7210-DE). Participants signed a written consent form before taking part in the study.

Full overnight PSG was recorded for clinical diagnosis of sleep apnea in two separate rooms. Additionally, an IR camera (Firefly MV, 0.3 MP, FMVU- 03MTM, FLIR Systems) was mounted about 1.4 m above the bed. The rooms were illuminated by IR light (Raytec RM25-F-50). Simultaneous and synchronized with PSG, IR videos were recorded from an overhead view with the aforementioned camera at a resolution of 640×480 at 30 frames per second. The 680×480 resolution captured the upper body part (the head, chest, and abdomen) with sufficient detail.

OA and CA were annotated by 3 trained individuals based on the RIP signal of the chest, the abdomen, the sum of the movement of the chest and the abdomen, nasal pressure, and oxygen saturation, following the American Academy of Sleep Medicine guideline [25,26]. The first 2 experts annotated all data with the agreement of 80%. The third expert annotated the discrepancies.

Data Analysis

Recorded IR videos were down-sampled from 30 Hz to 2 Hz to decrease the computational cost. The breathing rate during sleep is between 12 and 25 breaths per minute, that is, below 0.5 Hz. Therefore, the reduced sampling frequency of 2 Hz still satisfied the Nyquist rate for the respiratory signal by a wide margin. As each apnea lasts a minimum of 10 seconds, the sampling rate of 2 frames per second translates to a minimum of 20 image frames per event.

Frame-to-frame movements were tracked using dense optical flow. FlowNet 2.0 [27], a convolutional neural network (CNN) model, was used for this purpose. Optical flow generates a 2-channel image comprising the *x* (side to side) and the *y* (up and down) movement vector of each pixel from one video frame to the next. A 3D-CNN model was trained to analyze the sequence of optical flow images during each apnea to distinguish between OAs and CAs. This 3D-CNN model processed the optical flow of the entire image without explicit knowledge of where the person was at every frame of the video. For comparison, another CNN architecture was used to analyze the optical flow sequences in only the chest and abdomen regions. This model was trained to distinguish OAs from CAs via a late fusion of two 3D-CNN branches processing chest and abdominal movements. Performance was evaluated when either manually marked or automatically detected chest and abdomen regions were used. The performance of these CNN architectures was compared with three baseline models, including autocorrelation, movement histograms, and 2-dimensional fast Fourier transform (2DFFT)-CNN, which were chosen as commonly used approaches for motion analysis. These models were not applied previously for distinguishing OAs from CAs. The first two baseline models (autocorrelation and movement histograms) operate on the time-series data of optical flow movements, whereas the third baseline model (2DFFT-CNN) first transforms the signal into the frequency domain and then applies a CNN model to the resulting 2DFFT images.

3D Convolutional Neural Network

Sleep apneas last a minimum of 10 seconds [28]. A 3D-CNN model was trained on a sliding window of 10 seconds (20 frames at 2 Hz), starting from 5 seconds before the start of the apnea until 5 seconds after its ending. The size of the input tensor to the 3D-CNN model is 640×480×2×20 (image size 640×480 pixels; number of channels: 2; number of frames: 20). The two channels are outputs of the optical flow image, indicating the changes in the *x* (side to side) and *y* (up and down) directions. Details of the 3D-CNN model are shown in Table 1. RMSProp was used to optimize a class-weighted cross-entropy loss. The Matthews correlation coefficient was used for early stopping. An initial value of 0.001 for the learning rate and 1000 epoch with the batch size of 32 were used. The total number of parameters in this network was 95,649 of which 95,393 were trainable, and 256 were nontrainable.

OAs are more frequent than CAs; therefore, to achieve a more balanced training set, different stride was using for the sliding window in OAs and CAs (stride of 1 second for CA and stride of 5 seconds for OA). In test time, a stride of 1 second was used, and voting (overall strides within an event) determined the estimated label (OA vs CA) for that event.

Table 1. Architecture of a 3D convolutional neural network used to distinguish obstructive apnea from central apnea.

Layer	Number of filters, n	Size/stride	Activation function	Output size
Input	N/A[a]	N/A	N/A	480×640×20×2
Average pool	N/A	25×25×1/20×20×1	N/A	23×31×20×2
Convolutional	8	2×2×1/1×1×1	Linear	22×30×20×8
Dropout	N/A	N/A	N/A	22×30×20×8
Convolutional	16	3×3×5/1×1×1	N/A	20×28×16×16
Max pool	N/A	8×8× /2×2×1	N/A	7×11×16×16
Batch normalization	N/A	N/A	Leaky Relu[b]	7×11×16×16
Convolutional	64	2×2×2/1×1×1	N/A	6×10×15×64
Batch normalization	N/A	N/A	Leaky Relu	6×10×15×64
Convolutional	32	4×4×1/1×1×1	N/A	3×7×15×32
Batch Normalization	N/A	N/A	Relu	3×7×15×32
Dropout	N/A	N/A	N/A	3×7×15×32
Convolutional	16	2×2× /1×1×1	N/A	2×6×15×16
Batch normalization	N/A	N/A	Relu	2×6×15×16
Flatten	N/A	N/A	N/A	2880
Fully connected	16	2880×16	N/A	16
Fully connected	4	16×4	N/A	4
Output layer	N/A	4×1	Sigmoid	1

[a]N/A: not applicable.

[b]ReLu: rectified linear unit.

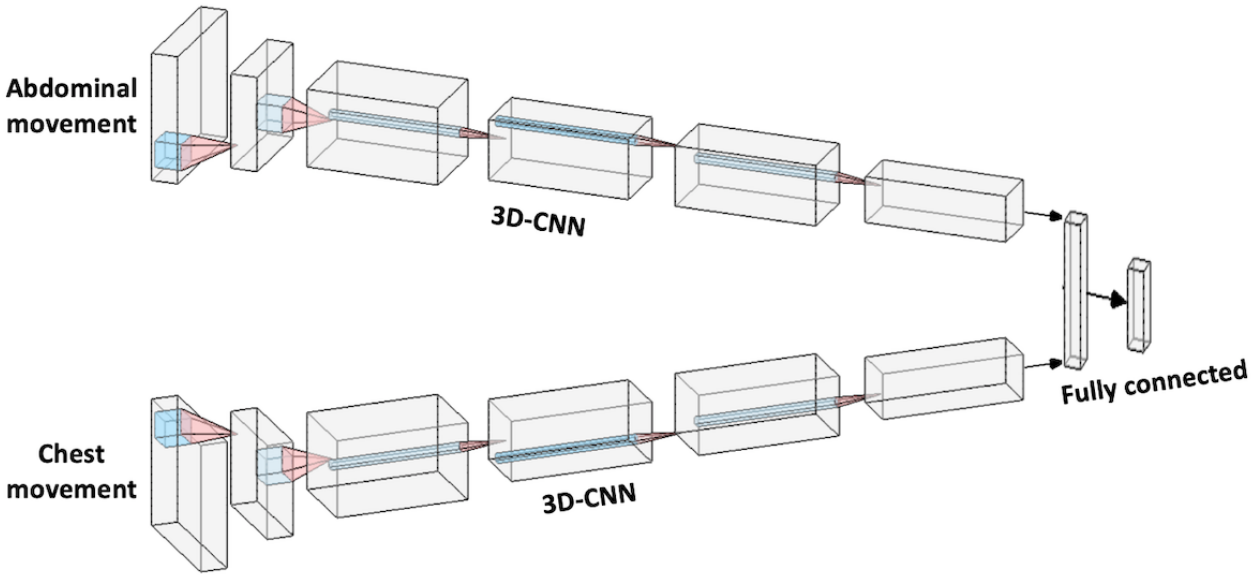
3D Convolutional Neural Network (Chest and Abdomen)

To investigate if knowledge about the location of the chest and abdomen at each image frame improves performance, the frames were manually marked as follows: for each person, at the first video frame of their sleep. After each position shift, a human annotator manually marked the locations of the chest and abdomen via two rectangles on the image. If the participant was covered by a blanket, the annotator used his/her best judgment to mark these locations (based on the current image frame and also by looking at previous and future image frames). Position shifts were automatically detected based on the total amount of movement in the scene, with a low threshold, so even small position shifts were not missed.

As the annotation of the chest and abdomen regions is subjective and time-consuming, we have also developed an algorithm to automatically find the chest and abdomen regions. To develop this model, as the chest and abdomen were often covered by a blanket sheet, it was challenging to train an object-detection CNN to detect them directly. Instead, a CNN model (YOLO v3) [29] was used first to locate the head, and another model was subsequently used to infer the position of the chest and abdomen bounding boxes. Specifically, a random forest regression model was trained based on the estimated head location, BMI, weight, height, head position (supine vs lateral), and body position (supine vs lateral) to estimate the bounding box of the chest and abdomen. The head and body positions were obtained automatically via a CNN-based model that was previously developed and validated [24].

Two 100×100 images were cropped around the estimated center of the chest and abdomen. Dense optical flow was computed in cropped regions of both the chest and abdomen. A 3D-CNN branch processed the sequence of chest movements, and another branch processed the sequence of abdominal movements. Outputs from both branches were concatenated into a fully connected network. The network was trained end to end. The architecture of the entire network is shown in Figure 2. The architecture of the 3D-CNN model in each branch is identical to the one shown in Table 1. The only difference is the first average pooling layer, which has the size and stride of 10 x 10 and 5 x 5, because of the smaller image size.

Figure 2. The Convolutional neural network architecture used to extract and combine information from movements of the chest and the abdomen. 3D-CNN: 3D convolutional neural network; CNN: convolutional neural network.

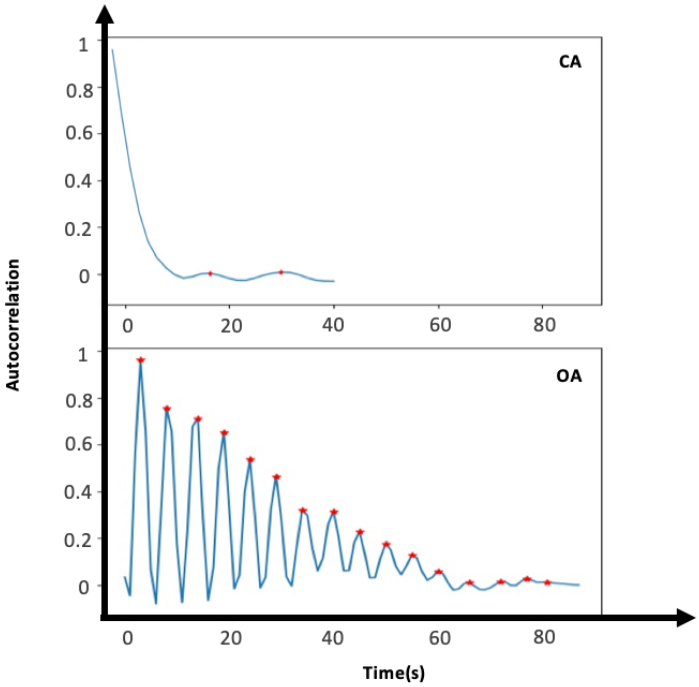


Baseline 1—Autocorrelation

This method uses autocorrelation to separate OAs from CAs based on their periodicity. The magnitude of the movement vector ($d=\sqrt{(x^2+y^2)}$) was calculated from the original 2-channel optical flow image (640×480×2) to form a single-channel movement magnitude image (640×480). Pixels with a large movement magnitude (>0.5 pixel per second), likely caused by large position shifts, were capped, that is, set to 0.5. The average of the movement image was then calculated for each event, leading to a 1-dimensional movement signal $m(t)$. A Butterworth band-pass filter with a lower cutoff frequency of 0.05 Hz and an upper cutoff frequency of 0.5 Hz was applied to the one-dimensional movement signal. Autocorrelation was computed for the filtered signals, and its first 10 peaks (if peaks did not exist, 0 was considered) were used to train three different binary classifiers to distinguish between OA and CA. Classifiers compared were linear support-vector machines, logistic regression, and random forest. Sample autocorrelation signals with their detected peaks are illustrated in Figure 3 for a CA and an OA.

is autocorrelation, where m_t is an event signal, is an average of signal, and is a shifted event with a lag of L. Lag (L) was set equal to the event duration. The summations are over all the values of t.

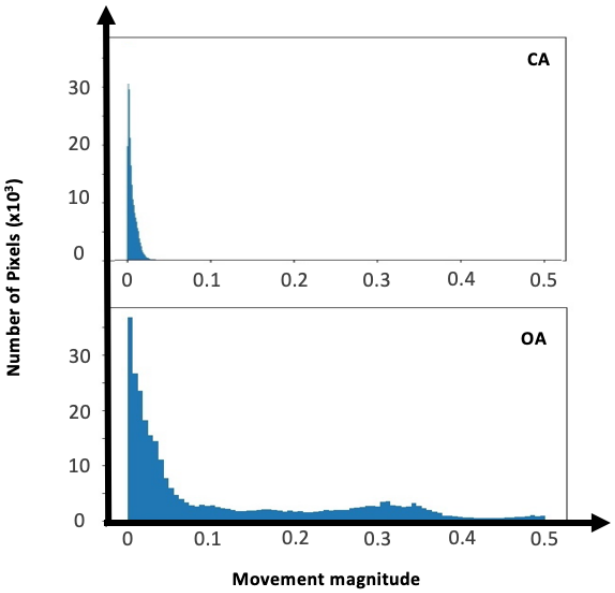
Figure 3. Autocorrelation signal of the movement in an obstructive apnea (OA) and central apnea (CA). OAs are more periodic due to the existence of breathing effort as compared with CAs. Therefore, the OA autocorrelation signal has more peaks, as indicated by red stars. CA: central apnea; OA: obstructive apnea.



Baseline 2—Movement Histograms

This method separates OAs from CAs based on their range of motion. Histogram of the movement magnitude was constructed for movements in the range of 0 to 0.5 pixels/second with a constant number of 1000 bins. Sample histogram signals for OA and CA are shown in Figure 4. The average of each bins (features) of histogram across an event was computed. Principal component analysis (PCA) was subsequently applied to reduce the number of bins. A random forest classifier was trained on the first 100 PCA components to distinguish OA from CA.

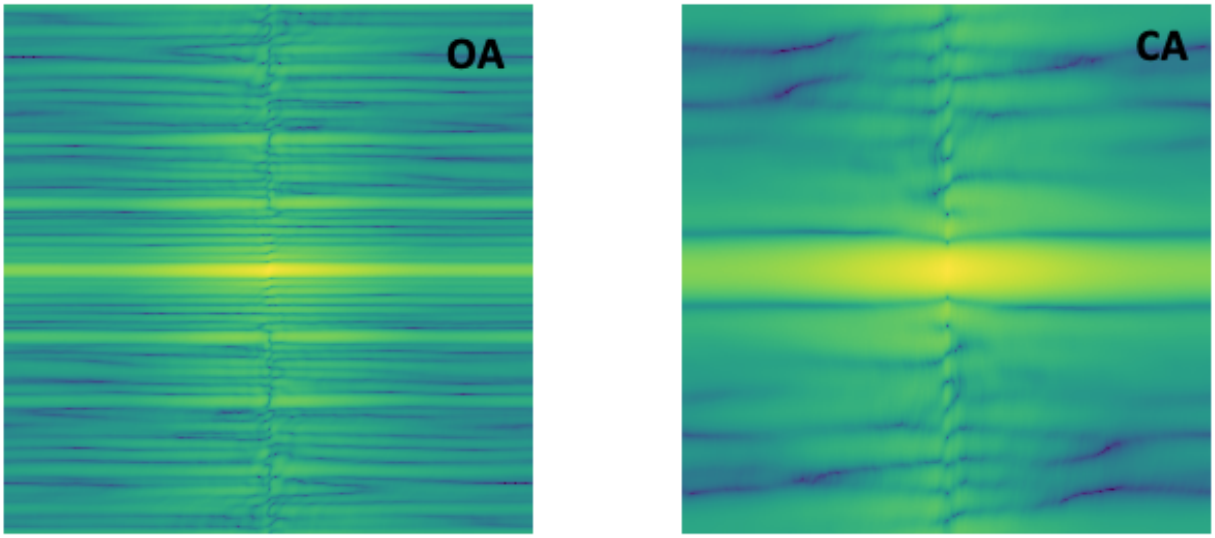
Figure 4. Histogram of movement magnitudes. Obstructive apneas have more range of motion as compared with central apneas because of the breathing effort. CA: central apnea; OA: obstructive apnea.



Baseline 3—2D Fast Fourier Transform-Convolutional Neural Network

This method separates OAs from CAs based on the frequency-domain representation of movement histograms. The movement histograms of each event were concatenated together over time to form an image and then transformed to the frequency domain via 2DFFT with a constant size of 128×128. A sample of 2DFFT signal is shown in Figure 5 for a CA and an OA. A CNN (DarkNet19) [30] was trained on the obtained 2DFFT image to distinguish between OA and CA.

Figure 5. 2D fast Fourier transform (2DFFT) of movement histograms for OA and CA. 2DFFT images of OA have a wider frequency range as compared with CA, as breathing effort during OA causes more fluctuation in the movement signal. 2DFTT: 2D fast Fourier transform; CA: central apnea; OA: obstructive apnea.



Validation

The data were divided into training and validation and test sets by the room in which the study was conducted to ensure the setup camera placement did not affect the algorithm performance. The recorded data of 21 participants (recorded in laboratory room number 1) were used in the training and validation sets. This set included 40 CAs and 313 OAs. The remaining 21 (recorded in laboratory room number 2) comprised the test set, which included 75 CAs and 299 OAs.

For the autocorrelation and movement histogram methods, classifier hyperparameters were tuned via 3-fold cross-validation on the training set. For the 3D-CNN (whole body) and 3D-CNN (chest and abdomen) methods, early stopping was based on performance on the validation set.

Performance of the head, chest, and abdomen bounding box detection was evaluated based on accuracy at the intersection over union values higher than 0.5 between predicted and manually annotated regions.

Results

Data from 42 participants (27 men and 15 women) were collected for this study. Participants' demographic information is shown in Table 2. None of the parameters were significantly different between the groups, except BMI with a *P* value of .04.

Figure 6 shows a sample image frame, as well as the manually marked and automatically detected bounding boxes, for the chest and the abdomen. The performance of the head, chest, and abdomen bounding box detection is quantified in Table 3. The head detection model obtained 92% accuracy, which was higher than the chest and abdomen detection models with accuracies of 83% and 67%, respectively.

Results of distinguishing OAs from CAs are shown in Table 4. Different classifiers obtained similar performance for the movement histograms method. For the sake of space, only results of the random forest classifier are shown in Table 4. The 3D-CNN model obtained the best performance with 95.4% accuracy and an F_1 score of 88.7%.

Table 2. Participant demographics (N=42).

Characteristics[a]	Room 1 (test set), mean (SD)	Room 2 (train set), mean (SD)
Male	8 (13)	7 (14)
Age (years)	53 (15)	55 (13)
BMI[b] (kg/m^2)	28 (6)	32 (7)
Sleep efficiency (%)	73 (18)	75 (18)
REM[c] sleep percentage (%)	16 (6)	15 (8)
Mean wake heart rate (bpm[d])	66 (17)	71 (15)
Mean REM heart rate (bpm)	63 (18)	72 (12)
Minimum SaO_2[e]	81 (9)	81 (7)
Mean SaO_2	94 (3)	94 (3)
Number of OAs[f] (events)	16 (35)	16 (23)
Number of CA[g] (events)	4 (10)	2 (3)
AHI[h] (events/hour)	24 (35)	29 (26)
Sleep duration (hour)	5 (1)	5 (1)

[a]Participants' information calculated from the sleep reports of the overnight sleep study of participants annotated by sleep technicians.

[b]BMI: body mass index. BMI is different between the two rooms with a *P* value of .04.

[c]REM: rapid eye movement.

[d]bpm: beats per minute.

[e]SaO_2: arterial oxygen saturation.

[f]OA: obstructive apnea.

[g]CA: central apnea.

[h]AHI: apnea-hypopnea index.

Figure 6. Sample chest, abdomen, and head detection results. Manually annotated and detected regions are shown in blue hashed line and orange solid line, respectively.

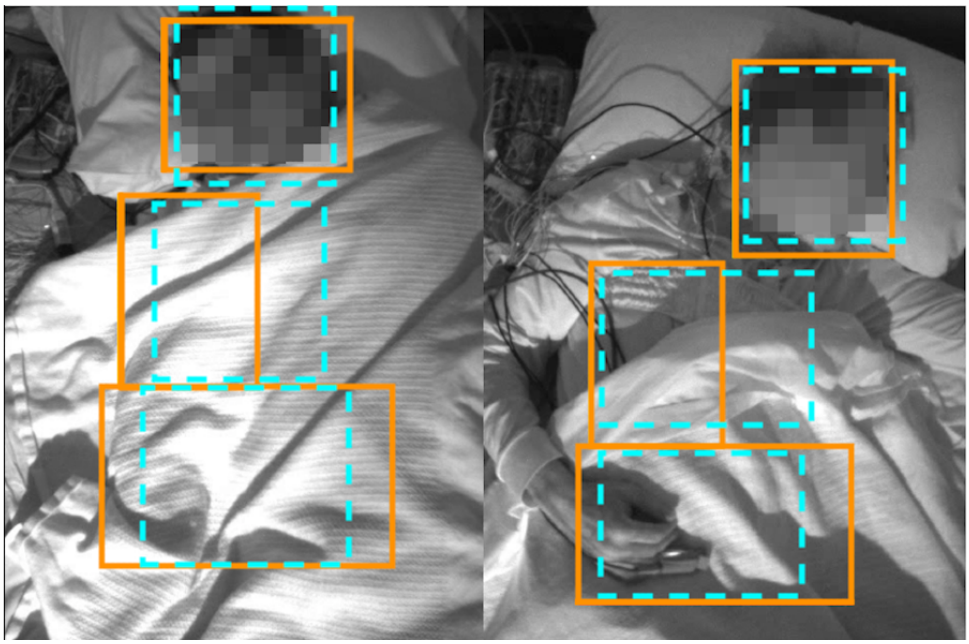


Table 3. Face, chest, and abdomen bounding box.

Detected object	Accuracy (%, at IoU[a] >0.5), mean (SD)
Head	92 (11)
Chest	83 (14)
Abdomen	67 (15)

[a]IoU: intersection over union.

Table 4. Obstructive apneas versus central apneas: prediction performance of different models.

Method	Accuracy[a] (%)	Precision[a] (%)	Recall[a] (%)	F_1 score[a] (%)
Autocorrelation	88.4	81.1	53.1	64.2
Histogram of movements	88.5	86.7	48.2	61.9
2DFFT-CNN[b]	89.7	69.1	75.6	72.3
3D-CNN[c]	95.4	88.2	89.3	88.7
3D-CNN chest and abdomen (annotated)	90.9	71.1	81.8	76.1
3D-CNN chest and abdomen (estimated)	89.3	72.1	76.0	74.0

[a]Accuracy, precision, recall, and F_1 score indicate the ratio of correct prediction to the total number of data points, the ratio of correct positive prediction to the total positive prediction, the ratio of correct positive prediction to the total positive data, and the harmonic mean of precision and recall.

[b]2DFFT-CNN: 2D fast Fourier transform-convolutional neural network

[c]3D-CNN: 3D convolutional neural network.

Discussion

Principal Findings

The proposed 3D-CNN model outperformed all three baseline methods. We hypothesized that localizing the chest and abdomen in the video will increase the signal-to-noise ratio to improve performance. However, as shown in Table 4, applying 3D-CNN on the whole image frame obtained the best performance. Figure 7 illustrates how the use of a blanket may explain these results. The blanket propagates chest and abdominal movements outside of their respective detected (or annotated) regions. As a result, localizing the chest and abdomen locations removed part of the respiratory-related movement signal. In addition, accurate detection of abdomen location is challenging when the body is covered by a blanket; however, increasing the number of data points (>42) could potentially address this challenge. Attention mechanisms could also potentially be used to automatically identify image regions in which chest or abdomen movements are prominent.

Figure 7. Annotated chest and abdomen regions do not capture a large area where most of the respiratory-related movement is visible. Manually annotated chest and abdomen regions are shown with blue boxes. Areas with large movement intensity (magnitude of the optical flow) are highlighted in pink.

Limitations

The methods developed and evaluated in this study are the first to use computer vision to differentiate between OAs and CAs. Although the models were externally validated on data collected in a different room, a limitation of this study was that the same camera model and camera setup was used to collect data in both rooms. Remedying this limitation will involve external validation on a dataset that will be recorded under different conditions, for example, use of another camera model, different viewing angle, or a different camera distance from the bed. This will evaluate how the models trained here generalize to potential variations that might occur in real-life scenarios, for example, in the home. Another limitation of the current system is that it relies on the assumption that apneas were already segmented, for example, using previously developed vision-based methods [21]. Remedying this limitation will involve evaluation of how the combination of such methods with the models developed here will perform.

Comparison With Prior Work

To address the challenges associated with PSG, there have been several investigations to develop convenient sleep apnea screening devices that can also distinguish CA from OA [14,15,28,31,32]. In a study proposed by Argod et al [15], pulse transit time technique was used to measure the delay between the R-wave on the electrocardiogram and a finger. They used the delay to visually classify CA from OA [15]. In another study, Park et al [31] designed an invasive implantable cardiac device to differentiate between CA and OA based on oscillation characteristics of the cardiac electrical activity. Luo et al [14] used the diaphragm electromyogram to track the activity of respiratory muscles to differentiate OA from CA. Thomas et al [28] used a single-lead electrocardiogram to classify OA from CA by measuring the elevated low-frequency coupling of heart rate variability and the fluctuations in the amplitude of the R-wave. These studies either are invasive or need the attachment of sensors to the body, which could be inconvenient and sensitive-to-motion artifacts and disrupt the user's regular sleep pattern. A noncontact method to distinguish CA from OA will address these challenges.

In one attempt, Nandakumar et al [32] tracked body movements via smartphones. They used frequency-modulated continuous-wave transmissions to find the motion changes in body. Although their model counted the number of CAs and OAs, the performance of the model on localizing and distinguishing of apneas was not reported. Moreover, this study did not report cross-validation results.

Conclusions and Future Works

This research project is the first vision-based noncontact method that differentiates sleep apneas by tracking body movements using IR videos. The developed algorithm was validated on 42 participants with various levels of sleep apnea severity. The algorithm performed well in distinguishing the OAs from the CAs. In this study, it was assumed that apneas were given. Future work can apply existing techniques [21] or novel CNN-based methods to identify apneas. Similar approaches can be used for noncontact assessment of respiration and respiratory effort during exercise testing in individuals who are using mechanical ventilators and infants with respiratory problems. Future work also involves collecting more data to perform an external validation of models developed here under varying conditions, as well as to improve chest and abdominal detection accuracy.

Acknowledgments

This work was supported in part by FedDev Ontario, in part by BresoTec Inc, in part by the Natural Sciences and Engineering Research Council of Canada through the Discovery Grant under Grant RGPIN 435653, and in part by the Toronto Rehabilitation Institute, University Health Network.

Conflicts of Interest

None declared.

References

1. Peppard PE, Young T, Barnet JH, Palta M, Hagen EW, Hla KM. Increased prevalence of sleep-disordered breathing in adults. Am J Epidemiol 2013 May 1;177(9):1006-1014 [FREE Full text] [doi: 10.1093/aje/kws342] [Medline: 23589584]
2. Kim NH. Obstructive sleep apnea and abnormal glucose metabolism. Diabetes Metab J 2012 Aug;36(4):268-272 [FREE Full text] [doi: 10.4093/dmj.2012.36.4.268] [Medline: 22950056]
3. Ayas N, Skomro R, Blackman A, Curren K, Fitzpatrick M, Fleetham J, et al. Obstructive sleep apnea and driving: A Canadian Thoracic Society and Canadian Sleep Society position paper. Can Respir J 2014;21(2):114-123 [FREE Full text] [doi: 10.1155/2014/357327] [Medline: 24724150]
4. Shahar E, Whitney CW, Redline S, Lee ET, Newman AB, Nieto FJ, et al. Sleep-disordered breathing and cardiovascular disease: cross-sectional results of the Sleep Heart Health Study. Am J Respir Crit Care Med 2001 Jan;163(1):19-25. [doi: 10.1164/ajrccm.163.1.2001008] [Medline: 11208620]
5. Arzt M, Young T, Finn L, Skatrud JB, Bradley TD. Association of sleep-disordered breathing and the occurrence of stroke. Am J Respir Crit Care Med 2005 Dec 1;172(11):1447-1451 [FREE Full text] [doi: 10.1164/rccm.200505-702OC] [Medline: 16141444]
6. Watson H, Sackner MA, Belsito AS. FreePatentsOnline. 1988. Method and Apparatus for Distinguishing Central Obstructive and Mixed Apneas by External Monitoring Devices Which Measure Rib Cage and Abdominal Compartmental Excursions During Respiration URL: http://www.freepatentsonline.com/4777962.html [accessed 2020-03-20]
7. Donovan LM, Kapur VK. Prevalence and characteristics of central compared to obstructive sleep apnea: analyses from the sleep heart health study cohort. Sleep 2016 Jul 1;39(7):1353-1359 [FREE Full text] [doi: 10.5665/sleep.5962] [Medline: 27166235]
8. Park JG, Ramar K, Olson EJ. Updates on definition, consequences, and management of obstructive sleep apnea. Mayo Clin Proc 2011 Jun;86(6):549-54; quiz 554 [FREE Full text] [doi: 10.4065/mcp.2010.0810] [Medline: 21628617]
9. Javaheri S, Dempsey JA. Central sleep apnea. Compr Physiol 2013 Jan;3(1):141-163. [doi: 10.1002/cphy.c110057] [Medline: 23720283]
10. Goldman MD. Interpretation of thoracoabdominal movements during breathing. Clin Sci (Lond) 1982 Jan;62(1):7-11. [doi: 10.1042/cs0620007] [Medline: 7035059]
11. Potts KJ, Butterfield DT, Sims P, Henderson M, Shames CB. Cost savings associated with an education campaign on the diagnosis and management of sleep-disordered breathing: a retrospective, claims-based US study. Popul Health Manag 2013 Feb;16(1):7-13. [doi: 10.1089/pop.2011.0102] [Medline: 22823227]
12. Javaheri S. CPAP should not be used for central sleep apnea in congestive heart failure patients. J Clin Sleep Med 2006 Oct 15;2(4):399-402. [doi: 10.5664/jcsm.26653] [Medline: 17557466]
13. Epstein LJ, Kristo D, Strollo PJ, Friedman N, Malhotra A, Patil SP, Adult Obstructive Sleep Apnea Task Force of the American Academy of Sleep Medicine. Clinical guideline for the evaluation, management and long-term care of obstructive sleep apnea in adults. J Clin Sleep Med 2009 Jun 15;5(3):263-276 [FREE Full text] [doi: 10.5664/jcsm.27497] [Medline: 19960649]
14. Luo YM, Tang J, Jolley C, Steier J, Zhong NS, Moxham J, et al. Distinguishing obstructive from central sleep apnea events: diaphragm electromyogram and esophageal pressure compared. Chest 2009 May;135(5):1133-1141. [doi: 10.1378/chest.08-1695] [Medline: 19118271]
15. Argod J, Pépin JL, Lévy P. Differentiating obstructive and central sleep respiratory events through pulse transit time. Am J Respir Crit Care Med 1998 Dec;158(6):1778-1783. [doi: 10.1164/ajrccm.158.6.9804157] [Medline: 9847267]
16. Glos M, Sabil A, Jelavic KS, Schöbel C, Fietze I, Penzel T. Characterization of respiratory events in obstructive sleep apnea using suprasternal pressure monitoring. J Clin Sleep Med 2018 Mar 15;14(3):359-369 [FREE Full text] [doi: 10.5664/jcsm.6978] [Medline: 29458696]
17. Flemons WW, Douglas NJ, Kuna ST, Rodenstein DO, Wheatley J. Access to diagnosis and treatment of patients with suspected sleep apnea. Am J Respir Crit Care Med 2004 Mar 15;169(6):668-672. [doi: 10.1164/rccm.200308-1124PP] [Medline: 15003950]
18. Li MH, Yadollahi A, Taati B. Noncontact vision-based cardiopulmonary monitoring in different sleeping positions. IEEE J Biomed Health Inform 2017 Sep;21(5):1367-1375. [doi: 10.1109/jbhi.2016.2567298]

19. Martinez M, Stiefelhagen R, editors. Breath Rate Monitoring During Sleep Using Near-IR Imagery and PCA. In: Proceedings of the 21st International Conference on Pattern Recognition. 2012 Presented at: ICPR'12; November 11-15, 2012; Tsukuba, Japan URL: https://ieeexplore.ieee.org/document/6460912
20. Zhu K, Li M, Akbarian S, Hafezi M, Yadollahi A, Taati B. Vision-based heart and respiratory rate monitoring during sleep - a validation study for the population at risk of sleep apnea. IEEE J Transl Eng Health Med 2019;7:1900708 [FREE Full text] [doi: 10.1109/JTEHM.2019.2946147] [Medline: 32166048]
21. Zhu K, Yadollahi A, Taati B. Non-Contact Apnea-Hypopnea Index Estimation using Near Infrared Video. In: Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2019 Presented at: EMBC'19; July 23-27, 2019; Berlin, Germany. [doi: 10.1109/embc.2019.8857711]
22. Deng F, Dong J, Wang X, Fang Y, Liu Y, Yu Z, et al. Design and implementation of a noncontact sleep monitoring system using infrared cameras and motion sensor. IEEE Trans Instrum Meas 2018 Jul;67(7):1555-1563. [doi: 10.1109/tim.2017.2779358]
23. Liao W, Yang C. Video-Based Activity and Movement Pattern Analysis in Overnight Sleep Studies. In: Proceedings of the 2008 19th International Conference on Pattern Recognition. 2008 Presented at: ICPR'08; December 8-11, 2008; Tampa, FL, USA. [doi: 10.1109/icpr.2008.4761635]
24. Akbarian S, Delfi G, Zhu K, Yadollahi A, Taati B. Automated non-contact detection of head and body positions during sleep. IEEE Access 2019;7:72826-72834. [doi: 10.1109/access.2019.2920025]
25. Kapur VK, Auckley DH, Chowdhuri S, Kuhlmann DC, Mehra R, Ramar K, et al. Clinical practice guideline for diagnostic testing for adult obstructive sleep apnea: an American Academy of Sleep Medicine clinical practice guideline. J Clin Sleep Med 2017 Mar 15;13(3):479-504 [FREE Full text] [doi: 10.5664/jcsm.6506] [Medline: 28162150]
26. Berry RB, Brooks R, Gamaldo C, Harding SM, Lloyd RM, Quan SF, et al. AASM Scoring Manual Updates for 2017 (Version 2.4). J Clin Sleep Med 2017 May 15;13(5):665-666 [FREE Full text] [doi: 10.5664/jcsm.6576] [Medline: 28416048]
27. Ilg E, Mayer N, Saikia T, Keuper M, Dosovitskiy A, Brox T. FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. 2017 Presented at: CVPR'17; July 21-26, 2017; Honolulu, HI, USA. [doi: 10.1109/cvpr.2017.179]
28. Thomas R, Mietus J, Peng C, Gilmartin G, Daly R, Goldberger A, et al. Differentiating obstructive from central and complex sleep apnea using an automated electrocardiogram-based method. Sleep 2007 Dec;30(12):1756-1769 [FREE Full text] [doi: 10.1093/sleep/30.12.1756] [Medline: 18246985]
29. Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv preprint 2018:- preprint. arXiv:1804.02767.
30. Joseph Redmon. 2013. Darknet: Open Source Neural Networks in C URL: https://pjreddie.com/darknet/ [accessed 2020-03-20]
31. Park E, Koh S. Google Patents. 2004. Differentiation of Central Sleep Apnea and Obstructive Sleep Apnea Using an Implantable Cardiac Device URL: https://patents.google.com/patent/US7225021B1/en [accessed 2020-03-20]
32. Nandakumar R, Gollakota S, Watson N. Contactless Sleep Apnea Detection on Smartphones. In: Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services.: ACM; 2015 Presented at: MobiSys'15; May 18 - 22, 2015; Florence, Italy p. 45-57. [doi: 10.1145/2742647.2742674]

Abbreviations

AHI: apnea-hypopnea index
CSA: central sleep apnea
CA: central apnea
CNN: convolutional neural network
CPAP: continuous positive airway pressure
2DFFT: 2-dimensional fast Fourier transform
IR: infrared
OSA: obstructive sleep apnea
OA: obstructive apnea
PCA: principal component analysis
PSG: polysomnography
RIP: respiratory inductance plethysmography

Edited by G Eysenbach; submitted 28.11.19; peer-reviewed by P Werner, R Stiefelhagen, M D'Souza, C Wu; comments to author 11.01.20; revised version received 24.02.20; accepted 25.02.20; published 22.05.20.

Please cite as:
Akbarian S, Montazeri Ghahjaverestan N, Yadollahi A, Taati B
Distinguishing Obstructive Versus Central Apneas in Infrared Video of Sleep Using Deep Learning: Validation Study
J Med Internet Res 2020;22(5):e17252
URL: http://www.jmir.org/2020/5/e17252/
doi:10.2196/17252
PMID:32441656

Original Paper

The Role of Health Concerns in Phishing Susceptibility: Survey Design Study

Mohamed Abdelhamid[1], PhD

Department of Information Systems, College of Business, California State University Long Beach, Long Beach, CA, United States

Corresponding Author:
Mohamed Abdelhamid, PhD
Department of Information Systems
College of Business
California State University Long Beach
1250 Bellflower Boulevard
Long Beach, CA, 90840
United States
Phone: 1 562 985 2361
Email: mohamed.abdelhamid@csulb.edu

Abstract

Background: Phishing is a cybercrime in which the attackers usually impersonate a trusted source. The attackers usually send an email that contains a link that allows them to steal the receiver's personal information. In the United States, phishing is the number one cybercrime by victim count according to the Federal Bureau of Investigation's 2019 internet crime report. Several studies investigated ways to increase awareness and improve employees' resistance to phishing attacks. However, in 2019, successful phishing attacks continued to rise at a high rate

Objective: The objective of this study was to investigate the influence of personality-based antecedents on phishing susceptibility in a health care context.

Methods: Survey data were collected from participants through Amazon Mechanical Turk to test a proposed conceptual model using structural equation modeling.

Results: A total of 200 participants took part. Health concerns, disposition to trust, and risk-taking propensity yielded higher phishing susceptibility. This highlights the important of personality-based factors in phishing attacks. In addition, females had a higher phishing susceptibility than male participants

Conclusions: While previous studies used health concerns as a motivator for contexts such as sharing personal health records with providers, this study shed light on the danger of higher health concerns in enabling the number one cybercrime.

(J Med Internet Res 2020;22(5):e18394) doi:10.2196/18394

KEYWORDS

phishing; health concerns; disposition to trust; risk-taking propensity; cybercrime; security, internet; trust; risk-taking; crime victims

Introduction

Background

Phishing refers to an internet cybercrime where a normal computer or mobile phone user is targeted by a cybercriminal through email. This communication is mostly intended to lure the user to provide their sensitive data such as personal health care records, passwords, bank information, and passwords. In a phishing attack, the attacker sends an email that impersonates a legitimate organization or person. The cybercriminal uses social engineering techniques to encourage the receiver to click on a suspicious link. The link can download a malicious app or provide a form that asks the recipient to enter sensitive personal information. Phishing attacks can target individuals, employees, corporations, or governments. Attackers are motivated by many factors such as achieving financial benefits or gaining a reputation in the cybercriminal community. An attacker can sell a stolen personal health care record for thousands of dollars [1].

According to the US Federal Bureau of Investigation's annual Internet Crime Complaint Center report for 2019 [2], there were about US $3.5 billion in financial losses alone due to cases of theft, fraud, and exploitation on the internet. The report states

that the most prevalent type of attack was the category Phishing/Vishing/Smishing/Pharming, various terms that are used to define different means to phish. For example, fraud conducted over the phone is termed *vishing*. *Smishing* refers to attacks conducted using texting. Phishing was ranked first by the number of victims, with more than twice as many victims as the second-ranked type of attacks. Phishing attacks have been increasing significantly in the past few years [3].

A recent IBM Security report ranked the health care industry first in term of the average cost of a data breach [4]. The report stated that phishing was one of the most common methods used in carrying out an attack. In 2018, about 15 million patients' records were breached in the United States. In addition, in the first 6 months of 2019, 25 million patients were compromised with phishing being the main factor behind most breaches [5].

Phishing not only has a negative influence on individuals' monetary assets but also builds doubt every single time they are contacted via phone calls, texts, or emails. For an individual, this shatters the reliability of electronic media to carry out a variety of tasks. This poses a great danger to digitizing traditional paper-based patient records.

According to various recent reports and research, the most common features of phishing emails are that they are astonishingly true, and have a sense of urgency (pushing the recipient to take an action as soon as possible), too many links, unexpected attachments, and an anonymous or unknown sender. Researchers and experts have been developing and testing ways of raising users' awareness so that they can detect phishing attacks.

However, efforts in raising awareness are yet to prove successful. As mentioned above, the success rate of the attacks has increased globally. About 35% of individuals do not even know what phishing is [6]. Many recent studies and current training focus on recommendations that many attackers get around. For example, Jensen et al [7] focused on training and recommendations that help individuals to avoid phishing. One of the main recommendations was to look for "https" in the address bar. Many online materials and articles list similar recommendations, including an unknown sender, generic greetings, and grammatical mistakes. However, according to a 2019 phishing trends and intelligence report, about 50% of phishing attacks use secure socket layer, making them harder to detect [8].

More importantly, attackers have constantly advanced and changed strategies. In fact, 2019 was a year of phishing evolution. Microsoft released a report in December 2019 that talked about evolving methods of phishing and explained the 3 most notable attack techniques of phishing they observed with their Microsoft Threat Protection services in 2019 [9].

Phishing is ultimately a social tactic. According the Verizon data breach investigation report [10], 43% of cyber attacks encompass social tactics, and of those that use social tactics, 93% are phishing attacks. Therefore, constant efforts should be invested in understanding the social and personal characteristics of individuals and victims. Detection, awareness, and training strategies need to be constantly evolving. In addition, one-size-fits-all recommendations and strategies will not benefit most individuals, organizations, or industries.

Literature Review

Research related to phishing can be categorized into two main streams. The first stream of research investigates technical aspects that can automatically detect phishing attacks using various methods such as machine learning and text mining. Jain and Richariya [11] proposed that a Web browser can also be trained to screen emails.

Methods such as text clustering, text mining, topic modelling, and classification have also been used to improve systems that detect and block phishing emails. For example, Basavaraju and Prabhakar [12] used cluster analysis to detect spam emails. Jeeva and Rajsingh [13] applied association rule mining techniques to detect phishing emails that contain malicious links.

Niakanlahiji et al [14] proposed a framework of machine learning to detect phishing webpages. The framework used 15 novel features that can be used on a webpage effectively without relying on search engines or other services. Some researchers have focused on improving text mining and data extraction techniques, which are then used as scripts to extract data from emails in a semiautomatic manner and analyze them to find patterns and other data [15].

However, in 2019, about 30% of phishing emails still bypassed security detection measures [16]. Phishing volume increased by approximately 41% in 2018, and the success rate of attacks has also increased. Although phishing detection technology is advancing, it cannot keep up with the advancement rate of phishing attacks. Therefore, many experts and researchers emphasize the importance of improving user awareness and training in regard to phishing. Most experts agree that the best way to defend against phishing attacks is to train employees and individuals to detect phishing emails in addition to security measures that detect some of the attacks automatically.

Therefore, the second stream of research has focused on the user side of the equation. User awareness of phishing emails stands to be one of the main prevention measures for phishing all around cyberspace. Hence, various awareness campaigns such as public announcements, seminars, and podcasts help to make users aware of these attacks and hence prevent phishing attacks in any form [17]. Miranda [18] suggested that phishing training programs can increase employee resistance to phishing attacks. Various studies proposed a game design framework to prevent many phishing attacks [3,19]. This framework enhances user avoidance behavior through motivation and hence results in preventing these phishing attacks. Other researchers looked at the characteristics of victims such as their email habits, perception of risk, and self-efficacy [20,21].

However, researchers have defined phishing as social engineering where an attacker attempts to fraudulently acquire sensitive information from a victim by impersonating a trustworthy third party. This brings in the scope for understanding normal users and their mentality gaps. Individuals are not always driven by rational thought and knowledge when making a decision or taking an action [22]. People may make

decisions quickly and may be driven by emotions or other factors [23]. For example, Jalali et al [24] found that perceived risk is not associated with click behavior, whereas workload is positively associated with click behavior. The limited number of studies looking at the influence of individuals' characteristics and traits on phishing susceptibility in a health care context creates a gap in the research.

Objective

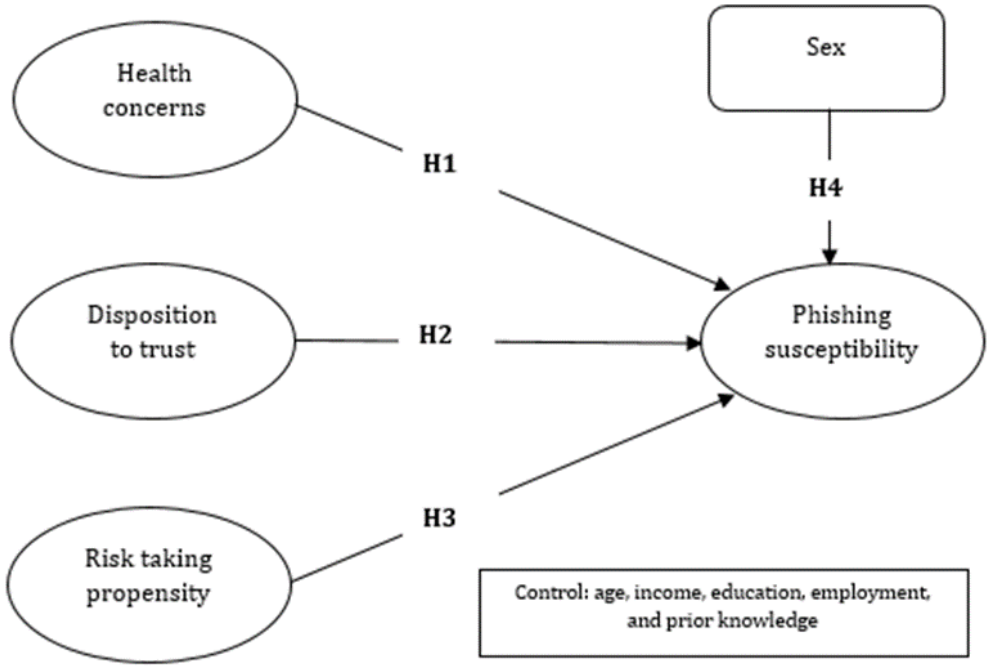
This study, motivated by improving phishing training and prevention measures by understanding individual characteristics, investigated the influence of personality-based factors—health concerns, disposition to trust, and risk taking—on phishing sustainability in health care–related attacks. The antecedents cover three areas of personality related to the context: concerns, risk, and trust. This allowed for identifying risk groups and understanding normal individuals and the mentality loopholes that an attacker uses to execute their tasks. This is one of the few studies that focused on personality-based antecedents as they relate to phishing susceptibility, specifically in the context of health care.

Methods

Conceptual Model

The following subsections argue for the relationships in the conceptual model shown in Figure 1. This is one the first studies that focused on personality-based variables and their influence on phishing susceptibility in the health care context.

Figure 1. Conceptual model.



Health Concerns

Health concerns refers to the degree to which an individual is concerned about their health [25]. Health-related traits have been shown to influence individuals' choices related to behaviors affecting their health directly or indirectly [26]. People who are concerned about their health are anticipated to pursue resources as they participate in behaviors to protect their health [25]. In the context of privacy, health concerns act as a promotor to seek better health outcomes [27]. In addition, the concept of health concerns is linked with a higher likelihood to seek health information online [28]. In general, Brelsford et al [29] found that patients with a high level of health concerns are motivated to take actions related to their health. However, the construct of health concerns has not been investigated in the context of phishing.

The first study hypothesis (H1, Figure 1) was that health concerns is positively associated with phishing susceptibility.

Disposition to Trust

Disposition to trust is a personality construct that refers to individuals' propensity to trust or distrust others [30]. In general, various types of trust have been extensively studied in different disciplines. The perception of trust can be linked with people or systems [31]. For example, the direct relationship between trust beliefs and use of technology has been well established [32-34]. The association between trust and information sharing has also been empirically demonstrated in previous studies [35-37]. These findings suggest that releasing personal information in exchange for e-services requires a great deal of trust. In the context of e-commerce and social media, prior literature has confirmed a positive association between trust in the system and a willingness to use the system [38,39].

The role of disposition to trust as a personality construct has been investigated in internet and website usage in general [30,40]. Bélanger and Carter [32] found that trust has an important role in promoting e-government use. Wang et al [41] investigated the effect of disposition to trust on mobile banking

adoption. In addition, disposition to trust has been linked to a higher likelihood of deception [42].

The second study hypothesis (H2, Figure 1) was that disposition to trust is positively associated with phishing susceptibility.

Risk-Taking Propensity

A key personality aspect that influences decision making is the individual's propensity to take on risks. An individual's risk-taking personality is defined as the behavioral propensity to seek rewards despite the probability of negative consequences [43]. This construct represents an attitude that is independent of any probability of outcome, but is anchored in how individuals value those outcomes [44]. Research has shown the link between different levels of risk-taking propensities and decision making in many contexts. Though the importance of individual differences in decision making has been examined in other fields, little information security research has investigated the impact of individual traits, especially in phishing susceptibility as it relates to health care. Hansen et al [45] found that risk-taking propensity has a direct positive influence on behavioral intention in the context social media use.

The third study hypothesis (H3, Figure 1) was that risk-taking propensity is positively associated with phishing susceptibility.

Role of Sex

Differences between males and females in their online health information sharing behavior is another problem worth investigating. Male-female differences in behavior with regard to phishing has been mostly investigated in general but not specifically in health care–related phishing scenarios. For example, Sun et al [46] found that male students score better than female students in antiphishing behavior. Similarly, Verkijika [47] reported consistent findings in a general phishing context. Results are expected to be consistent in the health care phishing scenario as well.

The fourth study hypothesis (H4, Figure 1) was that males have lower phishing susceptibility than females.

Data Collection

Data were collected online by administering the survey through Amazon Mechanical Turk (Amazon.com, Inc, Seattle, WA, USA), an online survey administration platform that allows for recruiting participants. The participants were redirected to take the survey built in Qualtrics^{XM} experience management software (Qualtrics). Participants had to be 18 years of age or older and reside in the United States. Many studies in the health care information technology context have used online data collection [27,48]. To measure phishing susceptibility (the dependent variable), participants were exposed to a scenario where they had to read an email and then indicate their intention to click on the email. The email was a phishing email in a health care context but participants were not told any information about its validity. The email was adopted from InfoSec [49].

Data Description

Data were cleaned and recoded using SAS version 9.4 (SAS Institute Inc). Responses that included missing data were removed from the final dataset.

Variables in the Model

All latent variables in the survey were borrowed from previous research and adapted to fit this study. Multimedia Appendix 1 shows the items and the source for variables in the model. All latent variable were measured using a Likert scale scored from 1 to 5. Sex was recoded to a binary variable named *Male* where 1 referred to a male participant and 0 referred to a female participant. Control variables were age, income, education, employment, and prior knowledge of the phishing concept.

Measurement Model

After the data were cleaned using SAS version 9.4, IBM SPSS Amos version 25 (IBM Corporation) was used to assess the latent validity and reliability of the latent variables and the overall fit of the measurement model. Confirmatory factor analysis was used to evaluate the overall measurement model.

The validity of the variables was tested using the average variance extracted (AVE). Finally, multicollinearity was assessed using variable inflation factor.

Results

Participant Characteristics

After the data cleaning process, a total of 200 valid response were included in the study. Table 1 shows the descriptive statistics of the participants. For example, 43% of the participants are female and 57% are male.

Table 1. Descriptive statistics of participants (N=200).

Variable, Category	Values, n (%)
Sex	
Female	86 (43.0)
Male	114 (57.0)
Age (years)	
18-25	25 (12.5)
26-35	113 (56.5)
36-45	31 (15.5)
46-55	19 (9.5)
>55	12 (6.0)
Education	
High school or less	23 (11.5)
Technical or community college	26 (13.0)
4-year college degree	105 (52.5)
Master's degree	42 (21.0)
Doctoral degree	2 (1.0)
Other	2 (1.0)
Income (US $)	
<25,000	41 (20.5)
25,000 to <50,000	74 (37.0)
50,000 to <75,000	58 (29.0)
≥75,000	27 (13.5)

Structural Equation Modeling

Structural equation modeling (SEM) was used to test the hypothesized model. Estimates derived from the SEM analysis were used to test the research hypotheses. The model explains 36.9% of the variance in phishing susceptibility. Figure 2 shows the SEM results for the hypothesized model.

Figure 2. Structural equation modeling (SEM) results. CFI: comparative fit index; RMSEA: root mean square error of approximation; TLI: Tucker-Lewis index.

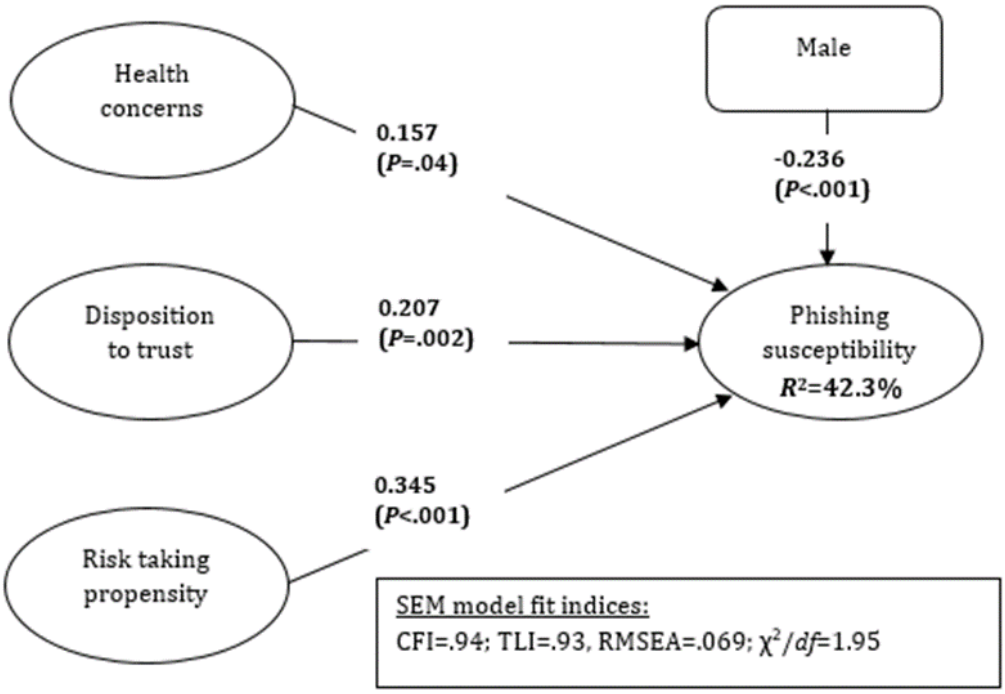


Table 2 show the results of the confirmatory factor analysis. The results show that the model fit well: root mean square error of approximation was .061, comparative fit index was .98, and Tucker-Lewis index was .97 [50]. In addition, all item loadings were high and significant, with scores ranging between .75 and .95.

All scores exceeded .5 AVE, which meets the cutoff for establishing convergent validity [51]. In fact, the smallest AVE score was .62. All AVE scores exceeded the squared construct intercorrelation for the corresponding variable, thus establishing discriminant validity [51]. All variables were reliable, as the construct reliability scores ranged from .83 to .96, exceeding the cutoff score of .7 for establishing reliability of the variable [52] (Table 2).

All variable inflation factor scores were well below the cutoff value of 10 [53]. Thus, there was no evidence to suggest the existence of multicollinearity. In addition, all variables were conceptually distinct.

Hypothesis 1 proposed a positive relationship between health concerns and phishing susceptibility. The SEM results supported this hypothesis. The path coefficient for the health concerns was positive and significant (β=.157, *P*<.04). These results suggested that health concerns lead to higher phishing susceptibility.

Hypothesis 2 suggested that disposition to trust leads to higher phishing susceptibility. The results supported this hypothesis (β=.14, *P*=.04). The path coefficients for disposition to trust were positive and significant, which provides evidence to support hypothesis 2. The finding proposes that individuals who have a propensity to trust others are more susceptible to phishing attacks. In addition, the magnitude of the path coefficient is higher than that of health concerns, indicating a larger influence of disposition to trust than of health concerns on phishing susceptibility.

Hypothesis 3 posited a positive relationship between risk-taking propensity and phishing susceptibility. The results show that the path coefficient was positive and significant (β=.345, *P*<.001), suggesting that risk-taking propensity yields a higher phishing susceptibility, thus supporting hypothesis 3. In addition, the magnitude of the path coefficient is the highest of all antecedents.

Hypothesis 4 argued that males, compared with females, have a lower phishing susceptibility. This hypothesis was supported by the results. The path coefficient for negative and significant (β=–.236, *P*<.001). This result is consistent with previous research [47].

Table 2. Confirmatory factor analysis results.

Variable, item[a]	Loading	Construct reliability	Average variance extracted	Variable inflation factor
Phishing susceptibility (PHS)		**.959**	**.887**	**N/A[b]**
PHS_1	.947			
PHS_2	.947			
PHS_3	.931			
Health concerns (HC)		**.832**	**.624**	**1.27**
HC_1	.805			
HC_2	.809			
HC_3	.754			
Disposition to trust (DTR)		**.89**	**.73**	**1.23**
DTR_1	.81			
DTR_2	.9			
DTR_3	.85			
Risk-taking propensity (RT)		**.904**	**.76**	**1.63**
RT_1	.832			
RT_2	.911			
RT_3	.87			
Measurement model goodness-of-fit indices				
RMSEA[c]	.061	N/A	N/A	N/A
Comparative fit index	.98	N/A	N/A	N/A
Tucker-Lewis index	.97	N/A	N/A	N/A
χ^2/df	1.737	N/A	N/A	N/A

[a]See Multimedia Appendix 1 for item details.

[b]N/A: not applicable.

[c]RMSEA: root mean square error of approximation.

Discussion

Principal Findings

This study investigated the relationship between personality-based factors and phishing susceptibility as it relates to health care. Personality-based factors have been shown to be associated with action, decision, attitude, and intention in various contexts [27,41,45]. However, limited research has focused on these relationship in a health care phishing context. This study focused on three main personality-based antecedents: health concerns, risk-taking propensity, and disposition to trust. This is one of the few studies that focused on these factors in the phishing context.

The study found that all three factors have in important role in leading to higher phishing susceptibility in a health care scenario. These findings suggest that personality-based factors should be taken into consideration when training individuals on phishing attacks and testing their phishing susceptibility and antiphishing behavior. The influence and effectiveness of training may differ based on personality-based traits. Thus, one-size-fits-all training, simulations, and strategy might not benefit most individuals.

While health concerns were found to be a motivator to engage patients in sharing their personal health information with providers [27], in a phishing scenario health concerns led to higher phishing susceptibility. Thus, health concerns could act as a "double-edged sword." Of all personality-based factors, the results suggested that risk-taking propensity had the highest effect on increasing phishing susceptibility. This implies that future research should study the high risk taker to find optimal methods to reduce phishing susceptibility. The study also confirmed previous findings that females are more susceptible to phishing attacks, but this research confirmed the finding in a health care context.

Limitations and Future Work

This study had several limitations, which can be addressed in future studies. The data were self-reported rather than actual behavior. Real-world behavior is very difficult to capture specifically in the health care context. In future studies, the influence of customized phishing training on individuals with different levels of personality-based factors will be tested.

Conclusions

Phishing remains a problem that continues to increase. While companies, experts, and researchers continue to develop new methods to detect phishing attacks and improve resistance to falling a victim to phishing, attackers are advancing and improving phishing attacks at a higher and a more successful rate. Microsoft named 2019 as a year of phishing evolution as attackers made innovations in both technical and social tactics. The only way to mitigate phishing attacks is get ahead of the attackers. Training and simulation should include a balance of customized material and approaches that fit the characteristics of the receiver.

This study contributes to the phishing literature by investigating personality-based factors and reporting findings that are new and important. In addition, the study contributes to the health care information technology literature by examining health care–related phishing scenarios and factors. There is limited research in the health care context that deals with phishing susceptibility.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Measurement.
[DOCX File , 164 KB - jmir_v22i5e18394_app1.docx]

References

1. Yao M. Your electronic medical records could be worth $1000 to hackers. Jersey City, NJ: Forbes Media LLC; 2017 Apr 14. URL: https://www.forbes.com/sites/mariyayao/2017/04/14/your-electronic-medical-records-can-be-worth-1000-to-hackers/#4252d5cb50cf [accessed 2020-04-13]
2. Internet Crime Complaint Center, Federal Bureau of Investigation. 2019 Internet Crime Report.: IC3; 2019. URL: https://pdf.ic3.gov/2019_IC3Report.pdf [accessed 2020-04-13]
3. Goel D, Jain AK. Mobile phishing attacks and defence mechanisms: state of art and open research challenges. Comput Secur 2018 Mar;73:519-544. [doi: 10.1016/j.cose.2017.12.006]
4. IBM Security. Cost of a Data Breach Report. Traverse City, MI: Ponemon Institute LLC; 2019. URL: https://www.ibm.com/downloads/cas/ZBZLY7KL [accessed 2020-04-13]
5. Davis J. Health IT Security. Danvers, MA: xtelligent Healthcare Media; 2019. The 10 biggest healthcare data breaches of 2019, so far URL: https://healthitsecurity.com/news/the-10-biggest-healthcare-data-breaches-of-2019-so-far [accessed 2020-04-13]
6. Wolinsky J. The cost of phishing emails to business is staggeringly high. Passaic, NJ: ValueWalk; 2019. URL: https://www.valuewalk.com/2019/07/cost-phishing-emails-infographic/ [accessed 2020-04-13]
7. Jensen ML, Dinger M, Wright RT, Thatcher JB. Training to mitigate phishing attacks using mindfulness techniques. J Manage Inf Syst 2017 Aug 17;34(2):597-626. [doi: 10.1080/07421222.2017.1334499]
8. PhishLabs. 2019 phishing trends and intelligence report: the growing social engineering threat. 2019. URL: https://info.phishlabs.com/hubfs/2019%20PTI%20Report/2019%20Phishing%20Trends%20and%20Intelligence%20Report.pdf [accessed 2020-04-13]
9. Office 365 Threat Research Team. The quiet evolution of phishing.: Microsoft; 2019 Dec 11. URL: https://www.microsoft.com/security/blog/2019/12/11/the-quiet-evolution-of-phishing/ [accessed 2020-01-09]
10. 2019 data breach investigations report. Basking Ridge, NJ: Verizon; 2019. URL: https://enterprise.verizon.com/resources/reports/2019-data-breach-investigations-report.pdf [accessed 2020-04-16]
11. Jain A, Richariya V. Implementing a web browser with phishing detection techniques. World Comput Sci Inf Technol J 2011;1(7):289-291 [FREE Full text]
12. Basavaraju M, Prabhakar DR. A novel method of spam mail detection using text based clustering approach. Int J Comput Appl 2010 Aug 10;5(4):15-25. [doi: 10.5120/906-1283]
13. Jeeva SC, Rajsingh EB. Intelligent phishing url detection using association rule mining. Hum Centric Comput Inf Sci 2016 Jul 10;6(1):10. [doi: 10.1186/s13673-016-0064-3]
14. Niakanlahiji A, Chu BT, Al-Shaer E. PhishMon: a machine learning framework for detecting phishing webpages. 2018 Presented at: IEEE International Conference on Intelligence and Security Informatics (ISI); Nov 8-10, 2018; Miami, FL, USA. [doi: 10.1109/isi.2018.8587410]
15. Zareapoor M, Seeja K. Text mining for phishing e-mail detection. In: Jain LC, Patnaik S, Ichalkaranje N, editors. Intelligent Computing, Communication and Devices. Cham, Switzerland: Springer Nature Switzerland; 2015:65-71.
16. Crane C. 20 phishing statistics to keep you from getting hooked in 2019. St Petersburg, FL: The SSL Store; 2019 Jul 24. URL: https://www.thesslstore.com/blog/20-phishing-statistics-to-keep-you-from-getting-hooked-in-2019/ [accessed 2020-04-13]

17. Jansson K, von Solms R. Phishing for phishing awareness. Behav Inf Technol 2013 Jun;32(6):584-593. [doi: 10.1080/0144929x.2011.632650]
18. Miranda MJ. Enhancing cybersecurity awareness training: a comprehensive phishing exercise approach. Int Manage Rev 2018;14(2):5-10 [FREE Full text]
19. Chandrasekaran M, Narayanan K, Upadhyaya S. Phishing email detection based on structural properties. 2006 Presented at: 9th Annual NYS Cyber Security Conference; Jun 14-15, 2006; Albany, NY, USA.
20. Vishwanath A, Harrison B, Ng YJ. Suspicion, cognition, and automaticity model of phishing susceptibility. Commun Res 2016 Feb 10;45(8):1146-1166. [doi: 10.1177/0093650215627483]
21. Arachchilage NAG, Love S, Beznosov K. Phishing threat avoidance behaviour: an empirical investigation. Comput Hum Behav 2016 Jul;60:185-197. [doi: 10.1016/j.chb.2016.02.065]
22. Soman D. The Last Mile: Creating Social and Economic Value From Behavioral Insights. Toronto, ON: University of Toronto Press; 2015.
23. Thaler R, Sunstein C. Nudge. London, UK: Penguin; 2009.
24. Jalali MS, Bruckes M, Westmattelmann D, Schewe G. Why employees (still) click on phishing links: investigation in hospitals. J Med Internet Res 2020 Jan 23;22(1):e16775 [FREE Full text] [doi: 10.2196/16775] [Medline: 32012071]
25. Jayanti RK, Burns AC. The antecedents of preventive health care behavior: an empirical study. J Acad Mark Sci 1998 Jan 01;26(1):6-15. [doi: 10.1177/0092070398261002]
26. Moorman C, Matulich E. A model of consumers' preventive health behaviors: the role of health motivation and health ability. J Consum Res 1993 Sep;20(2):208. [doi: 10.1086/209344]
27. Abdelhamid M. Greater patient health information control to improve the sustainability of health information exchanges. J Biomed Inform 2018 Jul;83:150-158 [FREE Full text] [doi: 10.1016/j.jbi.2018.06.002] [Medline: 29894811]
28. White RW, Horvitz E. From health search to healthcare: explorations of intention and utilization via query logs and user surveys. J Am Med Inform Assoc 2014;21(1):49-55 [FREE Full text] [doi: 10.1136/amiajnl-2012-001473] [Medline: 23666794]
29. Brelsford KM, Spratt SE, Beskow LM. Research use of electronic health records: patients' perspectives on contact by researchers. J Am Med Inform Assoc 2018 Sep 01;25(9):1122-1129 [FREE Full text] [doi: 10.1093/jamia/ocy087] [Medline: 29986107]
30. Wu G, Hu X, Wu Y. Effects of perceived interactivity, perceived web assurance and disposition to trust on initial online trust. J Comput Mediat Commun 2010 Oct 1;16(1):1-26. [doi: 10.1111/j.1083-6101.2010.01528.x]
31. Gefen D, Karahanna E, Straub DW. Trust and TAM in online shopping: an integrated model. MIS Q 2003;27(1):51. [doi: 10.2307/30036519]
32. Bélanger F, Carter L. Trust and risk in e-government adoption. J Strategic Inf Syst 2008 Jun;17(2):165-176. [doi: 10.1016/j.jsis.2007.12.002]
33. Posey C, Roberts TL, Lowry PB, Bennett RJ, Courtney JF. Insiders' protection of organizational information assets: development of a systematics-based taxonomy and theory of diversity for protection-motivated behaviors. MIS Q 2013 Apr 4;37(4):1189-1210. [doi: 10.25300/misq/2013/37.4.09]
34. Hung SY, Chang CM, Yu TJ. Determinants of user acceptance of the e-government services: the case of online tax filing and payment system. Gov Inf Q 2006 Jan;23(1):97-122. [doi: 10.1016/j.giq.2005.11.005]
35. Mou J, Shin D, Cohen JF. Trust and risk in consumer acceptance of e-services. Electron Commerce Res 2015 Nov 5;17(2):255-288. [doi: 10.1007/s10660-015-9205-4]
36. Anderson CL, Agarwal R. The digitization of healthcare: boundary risks, emotion, and consumer willingness to disclose personal health information. Inf Syst Res 2011 Sep;22(3):469-490. [doi: 10.1287/isre.1100.0335]
37. Abdelhamid M, Gaia J, Sanders GL. Putting the focus back on the patient: how privacy concerns affect personal health information sharing intentions. J Med Internet Res 2017 Sep 13;19(9):e169 [FREE Full text] [doi: 10.2196/jmir.6877] [Medline: 28903895]
38. Pavlou PA, Fygenson M. Understanding and predicting electronic commerce adoption: an extension of the theory of planned behavior. MIS Q 2006;30(1):115. [doi: 10.2307/25148720]
39. Lin WY, Zhang X, Song H, Omori K. Health information seeking in the Web 2.0 age: trust in social media, uncertainty reduction, and self-disclosure. Comput Hum Behav 2016 Mar;56:289-294. [doi: 10.1016/j.chb.2015.11.055]
40. McKnight DH, Choudhury V, Kacmar C. Developing and validating trust measures for e-commerce: an integrative typology. Inf Syst Res 2002 Sep;13(3):334-359. [doi: 10.1287/isre.13.3.334.81]
41. Wang SW, Ngamsiriudom W, Hsieh C. Trust disposition, trust antecedents, trust, and behavioral intention. Serv Ind J 2015 Jun;35(10):555-572. [doi: 10.1080/02642069.2015.1047827]
42. Wright RT, Marett K. The influence of experiential and dispositional factors in phishing: an empirical investigation of the deceived. J Manage Inf Syst 2014 Dec 08;27(1):273-303. [doi: 10.2753/mis0742-1222270111]
43. Gorka SM, Liu H, Klein D, Daughters SB, Shankman SA. Is risk-taking propensity a familial vulnerability factor for alcohol use? An examination in two independent samples. J Psychiatr Res 2015 Sep;68:54-60 [FREE Full text] [doi: 10.1016/j.jpsychires.2015.05.019] [Medline: 26228401]

44. Simonsohn U. Direct risk aversion: evidence from risky prospects valued below their worst outcome. Psychol Sci 2009 Jun;20(6):686-692. [doi: 10.1111/j.1467-9280.2009.02349.x] [Medline: 19422629]
45. Hansen JM, Saridakis G, Benson V. Risk, trust, and the interaction of perceived ease of use and behavioral control in predicting consumers' use of social media for transactions. Comput Hum Behav 2018 Mar;80:197-206. [doi: 10.1016/j.chb.2017.11.010]
46. Sun JC, Yu S, Lin SS, Tseng S. The mediating effect of anti-phishing self-efficacy between college students' internet self-efficacy and anti-phishing behavior and gender difference. Comput Hum Behav 2016 Jun;59:249-257. [doi: 10.1016/j.chb.2016.02.004]
47. Verkijika SF. "If you know what to do, will you take action to avoid mobile phishing attacks": self-efficacy, anticipated regret, and gender. Comput Hum Behav 2019 Dec;101:286-296. [doi: 10.1016/j.chb.2019.07.034]
48. Angst CM, Agarwal R. Adoption of electronic health records in the presence of privacy concerns: the elaboration likelihood model and individual persuasion. MIS Q 2009;33(2):339. [doi: 10.2307/20650295]
49. Dalasta D. Healthcare phishing attacks. Madison, WI: InfoSec; 2020. URL: https://resources.infosecinstitute.com/category/enterprise/phishing/the-phishing-landscape/phishing-attacks-by-demographic/phishing-in-healthcare/#gref [accessed 2020-04-13]
50. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equation Modeling 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
51. Fornell C, Larcker DF. Structural equation models with unobservable variables and measurement error: algebra and statistics. J Mark Res 2018 Nov 28;18(3):382-388. [doi: 10.1177/002224378101800313]
52. Joreskog K, Sorbom D. LISREL 8 User's Reference Guide. Mooresville, IN: Scientific Software; 1989.
53. Tabachnick B, Fidell L. Using Multivariate Statistics. Boston, MA: Pearson Higher Education; 2013.

Abbreviations

AVE: average variance extracted
SEM: structural equation modeling

Edited by G Eysenbach; submitted 24.02.20; peer-reviewed by S Guness; comments to author 16.03.20; revised version received 29.03.20; accepted 30.03.20; published 04.05.20.

Please cite as:
Abdelhamid M
The Role of Health Concerns in Phishing Susceptibility: Survey Design Study
J Med Internet Res 2020;22(5):e18394
URL: https://www.jmir.org/2020/5/e18394
doi:10.2196/18394
PMID:32364511

Review

Influences on the Uptake of and Engagement With Health and Well-Being Smartphone Apps: Systematic Review

Dorothy Szinay[1], MSc; Andy Jones[2], PhD; Tim Chadborn[3], PhD; Jamie Brown[4], PhD; Felix Naughton[1], PhD

[1]School of Health Sciences, University of East Anglia, Norwich, United Kingdom

[2]Norwich Medical School, University of East Anglia, Norwich, United Kingdom

[3]Behavioural Insights, Public Health England, London, United Kingdom

[4]Department of Behavioural Science and Health, University College London, London, United Kingdom

Corresponding Author:
Dorothy Szinay, MSc
School of Health Sciences
University of East Anglia
Norwich Research Park
Norwich, NR47TJ
United Kingdom
Phone: 44 1603593064 ext 3064
Email: d.szinay@uea.ac.uk

Abstract

Background: The public health impact of health and well-being digital interventions is dependent upon sufficient real-world uptake and engagement. Uptake is currently largely dependent on popularity indicators (eg, ranking and user ratings on app stores), which may not correspond with effectiveness, and rapid disengagement is common. Therefore, there is an urgent need to identify factors that influence uptake and engagement with health and well-being apps to inform new approaches that promote the effective use of such tools.

Objective: This review aimed to understand what is known about influences on the uptake of and engagement with health and well-being smartphone apps among adults.

Methods: We conducted a systematic review of quantitative, qualitative, and mixed methods studies. Studies conducted on adults were included if they focused on health and well-being smartphone apps reporting on uptake and engagement behavior. Studies identified through a systematic search in Medical Literature Analysis and Retrieval System Online, or MEDLARS Online (MEDLINE), EMBASE, Cumulative Index to Nursing and Allied Health Literature (CINAHL), PsychINFO, Scopus, Cochrane library databases, DataBase systems and Logic Programming (DBLP), and Association for Computing Machinery (ACM) Digital library were screened, with a proportion screened independently by 2 authors. Data synthesis and interpretation were undertaken using a deductive iterative process. External validity checking was undertaken by an independent researcher. A narrative synthesis of the findings was structured around the components of the capability, opportunity, motivation, behavior change model and the theoretical domains framework (TDF).

Results: Of the 7640 identified studies, 41 were included in the review. Factors related to uptake (U), engagement (E), or both (B) were identified. Under *capability*, the main factors identified were app literacy skills (B), app awareness (U), available user guidance (B), health information (E), statistical information on progress (E), well-designed reminders (E), features to reduce cognitive load (E), and self-monitoring features (E). Availability at low cost (U), positive tone, and personalization (E) were identified as physical *opportunity* factors, whereas recommendations for health and well-being apps (U), embedded health professional support (E), and social networking (E) possibilities were social *opportunity* factors. Finally, the *motivation* factors included positive feedback (E), available rewards (E), goal setting (E), and the perceived utility of the app (E).

Conclusions: Across a wide range of populations and behaviors, 26 factors relating to capability, opportunity, and motivation appear to influence the uptake of and engagement with health and well-being smartphone apps. Our recommendations may help app developers, health app portal developers, and policy makers in the optimization of health and well-being apps.

(J Med Internet Res 2020;22(5):e17572) doi:10.2196/17572

KEYWORDS

mHealth; health app; engagement; uptake; systematic review; COM-B; TDF; digital health; mobile phone; smartphone; smartphone app

Introduction

Background

Digital behavior change interventions, such as smartphone apps, can be effective and cost-effective tools to change a range of health-related behaviors [1,2]. For example, there have been promising studies of apps, including (1) delivering health prevention messages for men who have sex with men [3], (2) self-managing diabetes [4] and cardiovascular diseases [5], (3) weight management [6-8], (4) alcohol reduction [9-11], (5) mental health interventions [12], and (6) managing long-term conditions [13]. For certain behaviors such as reduction of alcohol consumption, they could also address the barriers experienced by health professionals when delivering brief interventions in person, such as lack of necessary training [11] and to reduce the stigma associated with alcohol consumption [2]. The public health implications are substantial because of their potential to have a low incremental cost and broad reach.

Despite their promise, effect sizes reported in evaluations of app-based interventions are often small. One potential explanation is the level of uptake and engagement. Uptake refers to the act of downloading and installing a smartphone app. Engagement has been defined as "(1) the extent (e.g. amount, frequency, duration, depth) of usage and (2) a subjective experience characterized by attention, interest and affect" [14]. To date, low uptake and poor engagement are commonly observed with digital interventions, which are often insufficient to sustain behavior change [15,16]. However, there is a lack of evidence regarding the main factors contributing to this problem.

Systematic reviews that focused on one specific behavior or a certain type of health or well-being app suggest that the effectiveness of evidence-based smartphone apps can be improved by targeting the design and engagement features, such as user-friendly design, individualized and culturally tailored content, or health professional support [17-19]. A review based on experiential and behavioral perspectives conceptualized key factors that might affect engagement with digital behavior change interventions: the content (eg, behavior change techniques, social support, and reminders) and how the content is delivered (eg, professional support, personalization, and aesthetic features) [14].

To our knowledge, no systematic review that primarily seeks to identify factors that influence the uptake of and engagement with a wide range of health and well-being smartphone apps has yet been conducted. To narrow the focus of this review, the four public health priority behaviors related to prevention (smoking, alcohol consumption, physical activity, and diet) along with mental health and well-being were targeted.

Theoretical Framework

The capability, opportunity, motivation, behavior (COM-B) model is a comprehensive framework that posits that individuals, to perform or change a behavior, need the capability to undertake it, the opportunity to take part in, and the motivation to engage with that behavior [20]. COM-B is increasingly being applied to inform the development of digital behavior change interventions [21-23]. The theoretical domains framework (TDF) [24] has previously been successfully applied for systematic reviews in other contexts [25,26]. The 14 domains of the TDF, described elsewhere [24], offer a concise coding framework that can be usefully conceptualized as possible targets for behavior change interventions. The TDF, being linked to the COM-B model [24], can be used as subthemes under the components of the COM-B model (see Multimedia Appendix 1).

Objectives

This systematic review aimed to synthesize factors identified in studies that influence the uptake of and engagement with health and well-being smartphone apps among adults targeting public health priority behaviors (smoking, alcohol consumption, physical activity, and diet) and mental health and well-being, and mapped these factors under the components of the COM-B model and constructs of the TDF. This could help inform stakeholders in public health and policy makers, digital behavior change intervention developers, and providers of health and well-being smartphone app portals to better target uptake and engagement.

Methods

Systematic Review

The review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA; Multimedia Appendix 2) [27], and the protocol was registered on the International Prospective Register of Systematic Reviews (CRD42019120312). The review used a mixed methods approach to generate different but complementary knowledge about users' views from qualitative findings and predictors and patterns of behavior from quantitative findings.

Eligibility Criteria

Eligible studies had to explore factors that influence uptake or engagement with health and well-being smartphone apps among adults. Table 1 summarizes the inclusion and exclusion criteria using the Population, Intervention, Comparison or Context, Outcomes, and Study Type tool.

Table 1. List of inclusion/exclusion criteria.

PICOS[a] component	Inclusion criteria	Exclusion criteria
Participants	• Adults ≥18 years. Studies including individuals ≥16 years were included if at least 70% of the participants were ≥18 years	Apps targeting health professionals
Intervention and context	• Studies investigating digital interventions using smartphone health and well-being behavior change apps on the following behaviors and outcomes: smoking, alcohol consumption, physical activity, diet and mental health, and well-being	Studies where the smartphone was not the primary intervention component
Outcomes	• Qualitative: findings described as facilitators, barriers, determinants of uptake, or engagement with health or well-being apps (either already existing or planned to be developed), including perceptions, beliefs, experiences, and interest of the participants. • Quantitative: uptake, measured as number of downloads, and engagement measured as number of logins, frequency of use, or any other relevant measure that tracks user engagement	Usability and user-testing studies, where functionality and app design were exclusively investigated for specific apps
Study design	• All study designs were included	None

[a]PICOS: Population, Intervention, Comparison or Context, Outcomes, and Study Type.

Search Strategy

Electronic Search

A systematic literature search was developed in consultation with a specialist librarian from the University of East Anglia and a senior information scientist from Public Health England (PHE). An iterative process helped to define the final search terms while ensuring a balance between sensitivity and specificity. A systematic literature search was performed in 8 electronic databases: Medical Literature Analysis and Retrieval System Online, or MEDLARS Online (MEDLINE), EMBASE, Cumulative Index to Nursing and Allied Health Literature (CINAHL), PsycINFO, Scopus, Cochrane library database, DataBase systems and Logic Programming (DBLP), and Association for Computing Machinery (ACM) Digital library. The databases were searched with no data limit, no publication or geographical restriction, but limited to the English language. Synonyms of 3 concepts were searched: (mhealth) AND (behavior change) AND (uptake or engagement; Multimedia Appendix 3 shows the MEDLINE search strategy). The electronic search was initially performed in November 2018 and was updated in August 2019.

Searching for Other Resources

Additionally, the search also included a manual search in key journals, such as *Journal of Medical Internet Research* and *Computers in Human Behavior*, and in *Google Scholar*. Reference lists of all included studies were hand-searched for additional studies. The search for gray literature included dissertations and theses, and unpublished research data and material were sought from government bodies and policy makers during stakeholder communication (PHE, National Health Service [NHS] in England).

Identification of Studies

All records identified by the search strategy were exported to Endnote X9 and deduplicated. To reduce the likelihood of reviewer selection bias and to assess how reliably the study eligibility criteria were applied, a subsample (10%) of records was additionally screened by a second reviewer (FN) during the title and abstract screening. Interrater reliability based on the number of eligible and ineligible studies was tested using Cohen's kappa statistics [28], with the following cut-offs being used: 0.41-0.60 to indicate moderate agreement, 0.61-0.80 substantial agreement, and 0.81-0.99 almost perfect agreement [28]. The full texts of potentially eligible studies were independently screened by DS, with 20% randomly selected and double-screened by FN. The exclusions of the studies were justified and recorded.

Data Extraction

A data extraction proforma was developed by the first author following the existing Cochrane guidelines [29], and the subsequent data were extracted: study characteristics (author, date of publication, sample size and type, location of the study, type of app investigated in the study, aim of the study, methodological characteristics such as design, data collection, and participants), main findings related to the research question of this systematic review (including participants' quotations and authors' interpretations in the qualitative studies and reported results of the quantitative studies), and conclusions of each study. The data extraction was performed by 1 reviewer (DS) and was checked for accuracy by a second reviewer (FN).

Quality Assessment

To assess the quality of the studies, critical appraisal was conducted using the latest version of the mixed methods appraisal tool (MMAT) [30]. MMAT is a unique tool [30] that was developed by pooling together the core relevant methodological criteria found in different well-known and widely used qualitative and quantitative critical appraisal tools [31-33].

The quality of all studies was assessed by the first reviewer (DS) and checked for accuracy by 2 other authors (FN and AJ). The tool is not intended to score the studies or to exclude papers but to offer a guide for interpreting findings [30].

Data Synthesis and Analysis

Integrative synthesis was applied to analyze the data [34,35]. The focus of the synthesis was on interpreting the data using specific concepts of the TDF as a deductive coding framework, which, for ease of interpretation, is summarized under the components of the COM-B model. Using the integrated approach, the data were pooled together by findings viewed as answering the same research questions, rather than by methods (eg, quantitative vs qualitative) [34,35].

Deductive thematic synthesis, a methodology designed to enhance the transparency of synthesizing qualitative data [36], was used to conduct the data synthesis of the findings of the qualitative studies and the qualitative component of the mixed methods studies. Using line-by-line coding, the findings were coded deductively into the domains of the TDF. The coding was conducted by the first author, and a randomly selected 10% of the coding was checked for accuracy by another author (FN). Regular coding meetings were conducted to maintain consistency. The expert opinion of an independent researcher with extensive experience in systematic reviewing was sought for data synthesis. The integrative approach includes interpretation of the quantitative findings by *qualitizing* [35], which refers to the textual interpretation of the findings of the quantitative studies (regardless of the interpretation of the author) so they can be combined narratively with qualitative data [35].

Results

Included Studies

A total of 7633 studies were initially retrieved, with a further 6 identified through manual search and reference check. An additional unpublished research report was received from stakeholders as part of the gray literature search process. No non-English papers were identified. A total of 2138 duplicates were removed. A total of 5429 studies were excluded based on the review of their titles and abstracts. Figure 1 illustrates the inclusion and exclusion of the studies following the guidance of the PRISMA flowchart [27].

During title and abstract screening, *substantial* agreement was achieved between the 2 independent reviewers (κ=0.63) [28]. Two types of disagreements were identified (one reviewer included studies that targeted app used in conjunction with a connected device and purely user-research studies) that limited agreement between the reviewers during the selection process, which were resolved through discussion and consultation with another author (AJ). After disagreements were resolved and the eligibility criteria were updated accordingly, 73 studies were identified as potentially meeting the inclusion criteria. All remaining titles and abstracts of records were assessed by 1 reviewer (DS). Of these, 41 studies were included in the review [37-77], out of which 13 were quantitative [41-44,49,53,55,63-65,68,76,77], 7 were mixed methods [38,47,59,62,73,74,78], and 21 were qualitative studies [37,39,40,45-47,50-52,54,56-58,60,61,66,67,70-72,75].

Description of the Included Studies

The end users of the studies were described as the general public [37,39,42,44,46,47,50-54,56-59,65,71,72,75,76], college students [48], existing app users [38,43,46,49,55,63,67,77,78], male workers in the male-dominated industry [60], lesbian, gay, bisexual, transgender, queer, and other spectrum of sexuality and gender (LGBTQ+) communities [40], rural communities [57], Asian ethnic minorities [41], pregnant women [73], patients in primary care [45,61,74], adult cancer survivors [62], adults with diabetes [57], those infected with HIV [64], those with chronic disease [68], and those with a bipolar disorder [69]. The focus of some studies was very specific and targeted a certain health behavior or condition, including alcohol reduction [38,46,54,58,59,64], smoking cessation [40,58,67,72,77], increasing physical activity [39,45,48,49,53,62,65,68], weight management [47,48,51,53,63,65,66,71,78], depression [52,61], mindfulness [50], diabetes management [57], and health management in pregnancy [73]. Other studies were less specific and targeted a more general mental health app [43,60,70] and a more general health app [37,41,42,44,55,56,74-76]. In all, 15 studies investigated factors influencing one particular app [38,39,43,45,46,49,50,54,55,63,65,67,70,72,77]. The remaining 27 studies examined users' perceptions of a wide range of apps or of a hypothetical app not yet developed.

The studies were published between 2011 and 2019 and were carried out in Australia [37,49,60,61,70], Belgium [69], Canada [40,51,55,67], China [68,73,76], Czech Republic [65], Ireland [45], Italy [39], New Zealand [47], Norway [75], Sweden [52], the United Kingdom [38,46,50,54,58,59,62,66,71,72,74], and the United States [41-44,48,53,56,57,63,64,77]. The study characteristics are summarized in Multimedia Appendix 4.

Quality Assessment of the Studies Included

On the basis of MMAT [30], the majority of the studies employing qualitative methodology were deemed to be of high quality. Concerns related to the sample were identified across many quantitative studies. This included issues around sampling and lack of clarity as to whether the groups were comparable at baseline or whether the sample was representative of the general population. In 4 nonrandomized studies, confounders were not accounted for by the design and analysis. Out of 7 mixed methods studies, 2 were judged to be of low quality, out of which one is an unpublished report (gray literature) and the other one is a published short report. See Multimedia Appendix 5 for details of the quality assessment for each study.

Data Analysis and Thematic Synthesis

Although not all the studies presented data for all aspects of this review, all studies presented some data that could be included in the synthesis. Evidence that was considered weakly explained or was judged to be unclear was not included in the summary of findings. An overview of the identified factors and the level of influence (uptake, engagement, or both) along with a brief description of each factor can be found in Table 2. Examples of supporting evidence are provided in the Textboxes 1-10.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart illustrating the inclusion and exclusion of studies.

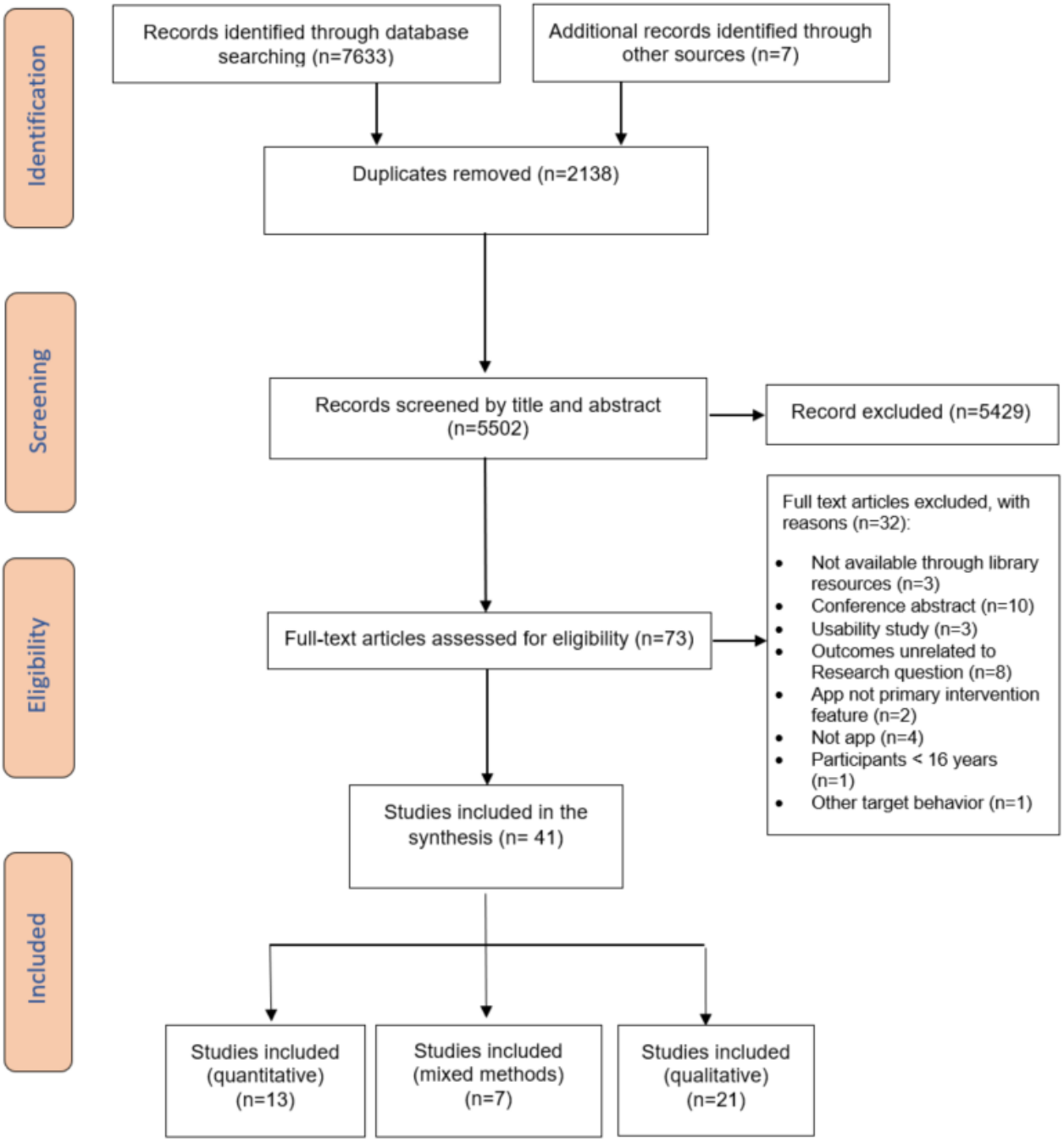


Table 2. Factors identified in the systematic review.

COM-B[a] component, TDF[b] construct, and identified factor (source)[c]	Uptake, engagement, or both	Short description of the factor
Physical capability		
Skills		
App literacy [46,50,57,61,65]	Both	Technological competency
Psychological capability		
Knowledge		
App awareness [54,56,57,61,75]	Uptake	Knowledge of the existence of health and well-being apps
User guidance [37,39,46,50,59,72]	Both	Instructions on how to effectively use the app
Health information [47,51,53,54,56-58,62,69,72,75,78]	Engagement	Educational information related to health and well-being aspects
Statistical information [37-39,46,52,54,57,66,67,71,72,75]	Engagement	A visual or numerical summary of progress
Memory, attention, and decision processes		
Well-designed reminders [37-40,43,46,48,51,52,54,56-58,62,66-69,71,78]	Engagement	The ability to customize reminders
Less cognitive load [37,39,46,48,50,51,54,56-58,60,66,69,71,72,75]	Engagement	The app is not too time consuming, easy to use, and requires minimal input
Coping games [40,60,67,72]	Engagement	Distraction activities within the app
Behavioral regulation		
Self-monitoring [36,38-40,45,48,51,52,55,57,59,60]	Engagement	The ability of the app to help self-regulation of the target behavior
Established routines [38,48,50,54,66]	Engagement	Regularity in using the app
Safety netting [37,61,66,73]	Engagement	Retaining the app for a potential precipitating event in the future
Physical opportunity		
Environmental context and resources		
Availability and accessibility [37,40,45,49,52,57,72,78]	Uptake	The ability to use a smartphone anytime anywhere
Low cost [37,40,47,48,56,68,72,74]	Uptake	The price of the app
Interactive and positive tone [46,51,57-60,69,71,72]	Engagement	Encouraging communication style
Personalization to needs [37,38,40,47,50,52,56,57,60-62,69,71,72,75,78]	Engagement	The possibility to use an app that is tailored to a user's needs
Social opportunity		
Social influences		
Recommendations [56-58,61,74]	Uptake	Suggestions received from other users
Health practitioner support [37,40,51,52,57,59,62,67,69,72,73]	Engagement	Possibility to get in touch with health professionals and practitioners within the app
Community networking [37,39,40,47,56,59,62,66-73,75]	Engagement	Social interaction with users with similar needs within the app or within their community
Social media [39,40,48,54,56,58,61,66,67,71,72,75,78]	Engagement	A choice to connect to social media platforms
Social competition [37,39,48,56,59,66,67]	Engagement	Competitive nature of the app with others or with themselves

COM-B[a] component, TDF[b] construct, and identified factor (source)[c]	Uptake, engagement, or both	Short description of the factor
Personification of the app [39,45,47,48,50,56]	Engagement	Applying human attributes to the app
Automatic motivation		
Reinforcement		
Feedback [37,39,45-48,51,52,54,56,58,62,67,72]	Engagement	Feedback regarding the user's performance
Rewards [37,40,45,46,56-59,66,69,71,75]	Engagement	Tangible and intangible reward in response to the user's effort
Emotions		
Curiosity [38,52,54,61]	Uptake	Desire to acquire knowledge and skills to use a behavior change tool
Reflective motivation		
Goals		
Goal setting [38,39,45,48,51,54,56,58,59,66,71,74]	Engagement	Establishing what the user would like to accomplish
Beliefs about consequences		
Perceived utility of the app [37,46,52,59,61,74]	Engagement	Discrepancy of what the users are looking for and what the app offers

[a]COM-B: capability, opportunity, motivation, behavior model.

[b]TDF: theoretical domains framework.

[c]Studies where the factors were identified.

Physical Capability

Theoretical Domains Framework: Skills

Skills refer to one's ability to perform an action and include constructs such as competencies, interpersonal skills, skill development, and practice (Textbox 1). App literacy [46,50,57,61,65], defined as technological competency to use a smartphone app, was reported by participants as being of high importance for both uptake and engagement. A basic level of app literacy is required to be able to download and initiate engagement with an app (see quote 1, Q1), whereas adequate app literacy skills would enhance users' intentions to engage with an app (Q2) [46,50]. In a cross-sectional study, advanced app literacy was associated with increased use of the social functions of an app, such as networking, but not with the functions that target action planning and goal management [65]. This suggests that app literacy might be an important aspect for successful uptake, but this alone might not be enough to maintain engagement. In contrast, users have reported that lack of app literacy skills could trigger negative emotions toward themselves (eg, self-blame and disappointment of not being able to use an app) [46,50,61] and could contribute to their perceived low self-confidence in using technology [61].

Textbox 1. Illustrative quotes (Q1 and Q2) for factors mapped onto the physical capability subcomponent of the capability, opportunity, motivation, behavior model and coded under the theoretical domains framework: skills.

Uptake and engagement

App literacy

- Quote 1: "I'd be happy to do it if I knew how to do it [but] I don't know how to download apps...I need help with technology. Like, I'm 58 and I didn't grow up in a technological age and so do find that I lack confidence with technology." [61]
- Quote 2: "I've never used it [these apps] because I never got it to work the way I wanted it to." [57]

Psychological Capability

Theoretical Domains Framework: Knowledge

Multiple factors were identified under the TDF domain that covers rational, procedural, and other types of knowledge; information; and awareness of the existence of something (Textbox 2). App awareness [54,56,57,61,75], such as information on the existence of health and well-being apps, would positively influence the uptake of health and well-being smartphone apps (Q3). It was suggested that many participants were not aware of the availability of such tools, and some found the disorganized nature of the commercial app stores confusing and represented a barrier for uptake [61].

Textbox 2. Illustrative quotes (Q3-Q13) for factors mapped onto the psychological capability subcomponent of the capability, opportunity, motivation, behavior model and coded under the theoretical domains framework: knowledge.

Uptake

App awareness

- Quote 3: "I didn't realize that they had an app." [57]

Engagement

User guidance

- Quote 4: "I want something to tell me 'Do number 1 first, then number 2. When you've done this go here' so I don't have to think too much about it. Once I've got it up and running I'm fine." [46]
- Quote 5: "Just at the beginning of the app, when you've downloaded it and you're using it for the first time, it should tell you what to do. But not every time. You don't need guidance how to use it and where things are, because I think it would just be annoying." [59]

Health information

- Quote 6: "[It is] important and really helps me to learn about bipolar disorder and read about stuff." [67]
- Quote 7: "I... enjoy learning something new. It's quite informative and makes you think about what you're doing. [QG] helps you to understand a bit more about what's going on...what could go wrong by continuing [to smoke]." [72]
- Quote 8: "I personally am scared of getting lymphedema, and still don't know sometimes what exercises are good to prevent it, so I think that maybe educating people about [...] consequences of not exercising from a really good NHS source would be helpful." [62]
- Quote 9: "I think everyone has heard that information many times. It's actually quite patronizing...shallow stuff, not hard-hitting useful facts. It obviously isn't a tailored app to each person, but it gives enough information that each person can relate to it in a tailored way. I find it really engaging, I suppose that's why I stuck with it." [72]

Statistical information

- Quote 10: "I like the numbers. I like to track stuff and have some figures behind it rather than just like, oh, I'll go for a run today. I'll be like, well, I'll go for a run today but what's my time from last time and how can I beat it? And I think that's why this kind of app appeals to me. If I just put the drinks in and it just said you're drinking too much but didn't give any numbers behind it, I'd probably delete it within a few days." [38]
- Quote 11: "It was like a visual of my day of smoking. And every day, you'd look at it, it went down and down and down, like it got better every day. So it was like a motivational thing to just look, like positive reinforcement." [67]
- Quote 12: "I couldn't find any graph that's reflected the mood so therefore I didn't see the point of having to fill that part out and I stopped filling it out." [46]
- Quote 13: "If you're having a bad day or a couple of bad days, seeing it on [the app] as a reflection [of your bad days] just like kicks you in the face even more, you know?" [67]

User guidance [37,39,46,50,59,72], namely, instructions on how to effectively use an app, such as how to create achievable goals, influenced uptake and initial engagement. It was proposed that having a guide on how to use an app could positively affect the users' intention to engage with it, and hence, users might be able to better regulate their behavior (Q4) [46,59]. However, the presence of a guide was reported off-putting and unnecessary for long-term engagement by producing negative emotions (eg, annoyance) once the knowledge regarding app functionality has been gathered (Q5) [59].

Available health information within the app was perceived by users as beneficial and positively influenced their engagement in several studies (Q6 and Q7) [47,51,53,54,56-58,62,69,72,75,78]. Depending on the target behavior, end users wished to (1) access advice on exercise routines [39,56,62,66]; (2) seek nutritional education [39,51,56,57,66,70]; (3) widen their knowledge of health consequences [58,67,72]; (4) find out more about healthy living while living with a medical condition [62,73]; (5) know more about the conditions they are living with [69,73,75]; (6) improve their health literacy [75]; (7) demystify myths [72]; (8) receive health news updates, such as on smoking taxes and bans [72]; and (9) better understand alcohol units in the UK [54].

However, the quality of information was identified as potentially affecting engagement [72]. Some users wanted a credible source, a trustworthy and evidence-based guide with references to the information they receive (Q8) [62,70,73]. Health information that focuses on negative aspects of past behavior that cannot be modified (eg, smoking or alcohol consumption) would trigger negative emotions (eg, regrets) [58]. It was suggested that better quality of information would increase the likelihood of maintaining users' engagement with an app, and consequently, they would better self-monitor their behavior [56,67]. This could be achieved by providing a wide range of information that everyone could relate to rather than facts that are already known (Q9) [72]. For example, 1 qualitative study suggested the use of health quizzes to promote engagement [75]. Health quizzes were also found promising by a large study that evaluated the uptake of a loyalty points–based health app conducted in Canada [55]. One of the intermediate objectives of that study was to

improve the Canadian population's health literacy by using health information related to quizzes. The app usage data included quiz completion rates, and the results showed that 60% of the users were highly engaged with the app by having more than 75% of health quizzes completed. Furthermore, better health literacy might enhance beliefs about consequences (eg, health outcome expectancies) [67,72] and the users' intention to stay engaged with an app and subsequently with the behavior they target to change [72,75]. Mackert et al [53] also found that adequate heath literacy was associated with increased engagement with fitness and nutrition apps.

Users valued available statistical information [37-39,46,52,54,57,66,67,71,72,75], which was a visual or numerical summary of progress or a trend in their behavior. This included features such as step counting [71,75], the number of calories consumed [54,71], number of days spent abstaining from smoking [67], the amount of money saved by quitting smoking [72] or by reducing drinking [54], a trend in their alcohol consumption and how it changes over time [38,46,54], as well as a way to allow analysis of user data [37,75]. Being able to check their progress helped users better monitor their behavior (Q10) [37-39,71,72], and for some individuals, a positive trajectory acted as a behavioral reinforcement (Q11) [46,67]. In 2 studies, participants reported that a lack of visual representation of progress led to disengagement with the alcohol reduction app (Q12) [38,46], and 1 study on smoking cessation reported negative emotions associated with progress viewing during *a few bad days*, suggesting discouragement (Q13) [67].

Theoretical Domains Framework Domain: Memory, Attention, and Decision Processes

This domain focuses on the ability to retain and select information, including aspects of attention, memory, decision making, and cognitive overload (Textbox 3). Reminders [37-40,43,46,48,51,52,54,56-58,62,66,67,69-71] to engage with an app were reported to be useful for people with busy schedules and for those who tend to forget engaging with the app and, therefore, with the target behavior [37,39,43,56,67]. Individuals described being inclined to check their phones when receiving a notification [37,38,40]. Reminders positively affected behavioral regulation by prompting engagement with self-monitoring and the tracking features of the app (Q14) [37,39,40,51,54,62,67,69-71] as well as reinforcing the users by reminding them about their positive progress [40,48,51]. A microrandomized trial found that a push notification that contained a tailored health message resulted in a small increase in the engagement with a health app [43]. A large study conducted on engagement with a weight loss app found that 16% of the most engaged group used reminders, compared with 1% of the least engaged group [64]. However, not all users found reminders useful [37,39,51,56-58,66]. In the case of behaviors that are associated with stigma (eg, alcohol consumption), reminders would threaten the users' social identity when they are received at an inappropriate time or wrong place (Q15) [38,46,54]. Therefore, the timing of when the reminders were sent as well as the language used appeared to be important conditions. If these conditions were not met, users were more likely to turn the notifications off [37,38,69] or ignore them (Q16) [56,66,67].

Textbox 3. Illustrative quotes (Q14-Q20) for factors mapped onto the psychological capability subcomponent of the capability, opportunity, motivation, behavior model and coded under the theoretical domains framework: memory, attention, and decision processes.

Engagement

Well-designed reminders

- Quote 14: "I found it was almost like having my girlfriend there, in a good way. So you're like, oh I haven't done this in two days, I didn't even realize, but my phone just reminded me. Better keep it going." [67]
- Quote 15: "I think because they were just pinging... and I was just thinking, I don't really want to read this right now. Obviously, and I don't know whether they do but I guess most people check their phone when something pings in and you can be with your friends and actually maybe you wouldn't want to be saying to your friends, I've just got a notification from Drinkaware." [38]
- Quote 16: "I completely ignored them [notifications]. Actually, I'm pretty sure I had the notifications that were from the app all turned off. It just felt like a pop up, like another thing for me to click close on throughout the day. I completely paid no attention to it." [67]

Less cognitive load

- Quote 17: "I really loved it [Couch to 5K], there was no excessive login, it was really easy you just downloaded and start you have to have your email, no password, no nothing like that, they don't send you a bunch emails that annoy the crap out of me. Nothing." [48]
- Quote 18: "What I'm thinking is, this better be easy, because otherwise I'm probably not going to do it. If there are too many obstacles in the way I won't. Even though I know I need to do this, I probably won't." [46]

Coping games

- Quote 19: "If there was a bunch of games on the app that were there to distract you from smoking, (you could) go play 5 mins of a quick game instead of smoking." [40]
- Quote 20: "Maybe if they had prior to like some type of like a mini game or something in there that would keep the mind occupied rather than telling you, "Don't smoke." [72]

Regarding attention and decision processes, the findings of the studies included in the review proposed that cognitive overload should be avoided to maintain engagement with an app. An app that is less time consuming, requires minimal input, and is easy

to use and log into was preferred (Q17) [37,39,46,48,50,51,54,56-58,60,66,69,71,72,75]. Additional functions that decrease the time spent on a task using an app were highly appreciated [37,39,48,50,54,56,71,72,75]. The automatization of data collection, for example, by linking apps to wearables [37] or by using the camera function to scan the barcodes to input calories [71] was found to be particularly useful in physical activity and weight management apps. An app that is easy to use and does not require extra effort would increase the intention to engage with it [39,46,48,54,56,57,74] and would improve users' self-monitoring and self-management strategies [48,51,66,75]. Conversely, using a difficult and time-consuming app would affect the users' perceived competence in engaging with it (Q18) [50]. Such an app often would be deleted or replaced with another app that is perceived to be easier to use [46,48,56,66,71]. Only 1 study found that users who are highly committed to change behavior (in this case, to reduce alcohol consumption) would be willing to overcome this barrier [54].

Including coping games [40,60,67,72] as distraction activities has been suggested as a helpful way to cope with cravings (smoking) [40,67,72] or with distress [60]. Some users indicated that by using their hands and minds, they expected to be preoccupied, instead of engaging with the undesirable behavior, while keeping them engaged with the app itself (Q19-Q20).

Theoretical Domains Framework Domain: Behavioral Regulation

Behavioral regulation refers to managing, monitoring, or changing actions or behavior (Textbox 4). Self-monitoring, the ability of an app to help monitor and regulate the target behavior [36,38-40,45,48,51,52,55,57,59,60], was found to be important in supporting behavior change. A self-monitoring feature was able to raise awareness about the number of cigarettes smoked [40,58], the amount of alcohol consumed [58], the number of steps taken [45], the mood they have [60], or users calorie intake (Q21) [48,56]. It also enhanced users' intention to engage with an app [51,52,58], provided *self-reinforcement* [52], helped increase self-efficacy (Q22) [56,61,71], and evoked feelings of *control, security, health, empowerment, and autonomy* [54].

An established routine or regularly using an app [38,48,50,54,66] positively affected the intention to engage with an app [50] and to maintain engagement (Q23). Furthermore, safety netting [37,61,66,73], defined as the ability of an app to provide *aftercare* [66] and an option to retain an app for a potential precipitating event in the future and for relapse prevention, was found to be useful to maintain the behavior, even when the target behavior has been achieved (Q24).

Textbox 4. Illustrative quotes (Q21-Q24) for factors mapped onto the psychological capability subcomponent of the capability, opportunity, motivation, behavior model and coded under the theoretical domains framework: behavioral regulation.

Engagement

Self-monitoring

- Quote 21: "You get a chance to see what you do on a daily basis, something you're probably not aware of." [56]
- Quote 22: "Because I can see I'm getting better, I use the app now, but I can see myself in the future not having to use it. Kind of like a stepping stone I guess." [71]

Routines

- Quote 23: "Because, I've got a couple of other little apps that I look at on a daily, not all apps, but a little regime of four or five, you know, I check the weather and I look at my drink app, and various things like that, a little routine, so pretty much daily." [38]

Safety netting

- Quote 24: "I think the migraine one's probably outlived its usefulness for me, but the back pain one, I could still go back to that at any time. If I started to need to monitor my pain again in a systematic way, I'd still go back to it." [37]

Physical Opportunity

Theoretical Domains Framework: Environmental Context and Resources

This domain refers to the circumstances of an individual's situation or environment that positively or negatively affects the uptake of or engagement with health and well-being smartphone apps (Textbox 5). The availability and accessibility of a smartphone [37,40,45,49,52,57,72,78] facilitate both uptake and engagement by having a behavior change device in close proximity (Q25). Although smartphones or tablets enhance the portability and accessibility of health apps, the development of an accompanying website was suggested to reduce inequality for those who might not have the opportunity to own a smartphone (Q26) [40]. Furthermore, the results of a digital behavior change intervention study examining engagement and nonusage attrition with a physical activity program suggest that when the app was used together with the accompanying website, a higher engagement rate was observed compared with those who used the app-only or the web-only versions [49].

The low cost of an app was found to be an influential factor for uptake [37,40,47,48,56,68,72,74] so that low-income individuals would be able to afford them (Q27) [47]. In a questionnaire study in China, 1 of the top barriers to using a health app was the extra cost, having a total of 83% of patients reporting that they would not be willing to pay for a health app [68]. Nevertheless, a few participants expressed their willingness to pay a small extra fee (ie, under US $5) if, this way, they could unlock unique features otherwise not available with the free version (Q28) [37,48,56,74].

Textbox 5. Illustrative quotes (Q25-Q34) for factors mapped onto the physical opportunity subcomponent of the capability, opportunity, motivation, behavior model and coded under the theoretical domains framework: environmental context and resources.

Uptake

Availability

- Quote 25: "It was real easy you just put it in your pocket and off you go and... you could do it at your own pace." [45]
- Quote 26: "I feel like there would need to be a website equivalent with it (for) people who don't have access to smartphones but do have access to public libraries. A lot of smokers are LGBTQ and a lot of LGBTQ are in poverty and homeless. The people that you want to access might not be able to access the program." [40]

Low cost

- Quote 27: "I wouldn't pay money for an app. I think that's kinda stupid." [48]
- Quote 28: "I'm prepared to pay for applications. As well as being in the software industry, I understand that it's people's livelihoods are attached to this. I use some free applications, but I often will pay for the upgraded or the purchased option." [37]

Engagement

Positive tone

- Quote 29: "I had a chocolate bar today and It would say, this chocolate bar contained this much saturated fat and... I just feel really guilty now." [71]
- Quote 30: "I think I'm more likely to listen to practical advice rather than finger wagging..." [58]
- Quote 31: "I just see it as a way to help me monitor what I'm doing and maybe give me a little kick in the pants every now again to be like, 'By the way, that donut had five hundred calories in it. Maybe make a better choice at dinner'." [51]

Personalization

- Quote 32: "The more I would be able to manipulate the app to be and do what I wanted or needed, for my own circumstances, the more likely I am to use it." [59]
- Quote 33: "It must be very personalized, it's easy to find things on the Internet, but it's mostly for normal people." [75]
- Quote 34: "Assuming that it's customised to LGBTQ (and) it incorporates the kinds of struggles that we've lived through, it wouldn't be any average quit-smoking app. The fact that it's specific to a community... the fact that it's LGBTQ-specific, that would help us more than if it was just a general quit-smoking app." [40]

Numerous studies have found that interactivity and positivity of tone may be efficacious for engagement, especially when attempting to change behaviors associated with self-blame (eg, weight management) (Q29) [46,51,57-60,69,71,72]. In total, 3 studies provided evidence that an encouraging tone rather than a condescending tone was important [46,58,69]. Evidence from 1 study suggested that apps should use praise but avoid shame [51], and another study provided evidence that a relaxed tone may be beneficial and may include jokes [46]. Several studies suggested that demanding or annoying language would be ignored (Q30) [57-59], although a study of nutrition apps reported the occasional need for a tougher attitude to achieve goals (Q31) [51]. Nevertheless, careful selection of the terminology used to understand the app and what it does, such as using simple and clear language, was suggested to make a noteworthy difference in the effectiveness of the content [60,72]. Terminology around certain behaviors might make a difference. For example, it was reported that using a *nonsmoker* label as opposed to an *ex-smoker* label would increase people's self-confidence [72]. It was suggested that unsupportive language would evoke negative emotions (eg, guilt and regret), which would affect the intention to engage with an app [46,59,71].

A personalized app was highly valued for engagement [37,38,40,47,50,52,56,57,60-62,69-72,75]. Users would want to have control over the app (Q32) [59,66,69]. They would like to be able to switch off features they do not use [37], and to use external incentives, such as uploaded photos or quotes [66,67], and to personalize their goal and how to achieve it [40]. Users would also like to choose a level where to start using a particular app. For example, a more experienced user would want to have the possibility to start a mindfulness practice at the intermediate level rather than at the beginner level [50]. Users were seeking to receive more personalized information about their current behavioral habits, demographic characteristics, long-term effects of the current behavior [38,56,60,78], and recommendations based on their tracked data [57]. Personalization can also be extended to their identity (Q33). Participants were looking for an app that is tailored to their cultural and social identities, such as LGBTQ+ people, cancer survivors, or other patients who are predisposed to have other struggles and mental health issues (Q34) [40]. Personalization to users' needs and preferences suggested better engagement [58,59,61], whereas lack of flexibility in content was found to be a reason to stop engagement [52], and in some cases, it created frustration [71]. Furthermore, a large study found that 30% of the most frequently engaged group customized the app more, for example, uploaded pictures, than the least engaged group (2%) [63].

Social Opportunity

Theoretical Domains Framework: Social Influences

Social influences are interpersonal influences (received from other individuals) that could impact an individual's behaviors, decisions, thoughts, and feelings (Textbox 6). In 5 studies, recommendations to use an app [56-58,61,74], received from health care practitioners or trusted providers [57,61,74], friends and families [56,60,74], or by reading user reviews [56,58,74], positively affected the uptake of health and well-being apps (Q35-Q37).

Connections between an app and health practitioner support were highly valued [37,40,51,52,57,59,62,67,69,72,73]. Participants reported that counseling services should be linked to an app [40,67,69], such as an *emergency button* feature [69], whereas others have emphasized the importance of linking an app to their health care provider (Q38-Q40) [37,62]. Health practitioner support could help overcome potential barriers caused by lack of skills, such as app literacy [52]; enhance self-monitoring [52,62]; and act as reinforcement [52], having the potential to enhance intentions to engage with the app (Q40) [52,62,72].

The possibility of community networking within apps with other users or other people with similar needs has been identified in multiple studies [37,39,40,47,56,59,62,66,67,69-73,75]. It was considered an important social support by reinforcing behavior change [47,56,59,62,69,72,73] and by sharing knowledge and experiences [37,69,73,75]. This was found to increase their intention to engage with the app and, subsequently, the behavior (Q41-Q42) [62]. A large study found that the most engaged group had a mean number of 24 friends within the app, as opposed to the least engaged group (1 friend) [64]. Users' potential social roles or group identities and personal preferences should be taken into consideration. For instance, individuals from the LGBTQ+ community [40] and cancer survivors [62] would wish to interact with people who face similar challenges (Q41). In addition, some users would not want to share information with strangers due to fear of social comparison [39,59] or social stigma [54], whereas others were more open to connecting with strangers rather than with friends or family (Q42-Q44) [56].

Evidence for the importance of embedded social media for engagement has been mixed [39,40,48,54,56,58,61,66,67,70-72,75]. It largely depends on the individual's attitude toward these channels and on the target behavior. Some users found this reinforcing (Q46) [40,61,71,75], whereas others did not want to engage with such features due to social stigma (eg, smoking, alcohol consumption, or weight management; Q46-Q47) [39,48,54,56,58,67,72].

Social competition [37,39,48,56,59,66,67] includes the possibility for individuals to compete with themselves (ie, their previous achievements or breaking their own records) or with other app users (Q48-Q49). A total of 5 studies suggest that the reinforcing nature of social competitions might increase the intention to engage with an app [37,48,56,59,66]. The increased engagement was anticipated when the competition is based on support by receiving encouragement from others [39,67], rather than on defeating each other, which might prompt discouragement to use the app (Q50) [67].

Several studies described that some participants felt that apps can impersonate a little person [39,45,47,48,50,56], which increased the intention to use the app (Q51-52) [45,48,50]. It was also suggested that if the app is too impersonal, it would not offer the social support the users' need [47]. In contrast, in 2 studies, the participants were concerned about having a machine telling them what to do (Q53) [47,56].

Moreover, personal experience related to noncommunicable diseases might increase the chances of the uptake of apps. One study conducted on Latino and Asian subgroups in the United States found that the odds of downloading a health app was twice as high for those who had a family history of heart attack (odds ratio 2.02, 95% CI 1.16-3.51), compared with those who did not [41].

Automatic Motivation

Theoretical Domains Framework: Reinforcement

Reinforcement is a process or action of encouraging a pattern of behavior (Textbox 7). Users reported better engagement when positive feedback was received (Q54) [37,39,45-48,51,52,54,56,58,62,67,72]. Visual feedback of progress made users aware of their advancement in reaching their goal (Q55) [37,45,46], whereas auditory feedback was seen as encouraging during physical activity (eg, running) [37,48]. For some, instant feedback on their progress, even if it is of a positive nature, was perceived to cause pressure and potential disappointment if they were not able to reach their goal (Q56) [45,56].

Offering rewards [37,40,45,46,56-59,66,69,71,75] was found to be a useful way to increase engagement. Participants suggested including gamification elements in apps to enhance engagement [37,56,69,71,75]. Some users found intangible rewards (eg, badges) motivating (Q57) [46,56,58,59,66,71], whereas others would want to receive tangible rewards instead (eg, free t-shirt, gift cards, cash, reduction in health insurance, or vouchers provided by hospitals or doctor's office; Q58-Q59) [40,56,58,66]. This has been partly supported by 2 quantitative studies. In 1 study, having a health insurance was associated with uptake of, but not with engagement with, health apps [42]. Another study found that when offering loyalty points, engagement increased for at least three months [55].

Textbox 6. Illustrative quotes (Q35-Q53) for factors mapped onto the social opportunity subcomponent of the capability, opportunity, motivation, behavior model and coded under the theoretical domains framework: social influences.

Uptake

Recommendations

- Quote 35: “I’d rather ask a counselor or a doctor what they would recommend.” [61]
- Quote 36: “Most of mine [my apps] are friend recommendations, people with similar activities.” [56]
- Quote 37: “...if an app has a good rating, despite the one or two people who are not satisfied, I think it would mean that it works for the majority of people.” [58]

Engagement

Health practitioner support

- Quote 38: “It would help in times of crisis to be able to be in touch with a professional, or if I needed to ask health questions related to alcoholism.” [59]
- Quote 39: “I want to let others know when I’m not well, the app would help me.” [69]
- Quote 40: “The therapist helped me to find my motivation every now and then, and then I was on top of it for about a week or so, and eventually the application sort of became a part of my everyday life. Then it was pretty obvious that I would use it and then I didn't even think about whether it was hard to use it, I just did it.” [52]

Community networking

- Quote 41: “It is so important to get in touch with people who went through the same thing as you have. [...] I think that if an app for cancer survivors had a forum on it as a part of the application to motivate each other, that would be amazing.” [62]
- Quote 42: “I don't think I would share on the social media, but within the app community I think it is important to like inspire and be motivated by others.” [66]
- Quote 43: “So having some sort of platform where everyone can just say, ‘This is how I stopped’ or ‘This is how I'm trying to stop’ and then other people giving feedback saying, ‘This is good’ or, ‘This is not’.” [72]
- Quote 44: “Being able to exchange feedback with strangers with the same goal could be supportive but non-judgemental as you will probably not know the other users.” [59]

Embedded social media

- Quote 45: “Integrating it with the social media is definitely a great thing to do because they can always fall back to Facebook, Twitter, etc. And through this, people can get to share their experiences and keep an update and tell whatever experiences they may have to share. So it’s like ongoing support.” [40]
- Quote 46: “Yeah you can share on Facebook and stuff, but I hate that. I hate when apps sync to like every form of social media. I’m like really weird about social media, so, no I don’t want to share it.” [48]
- Quote 47: “Don't want to share progress on social media in case you fail.” [72]

Social competition

- Quote 48: “Whenever we do a weekend challenge, you always have a look at what the other person's doing and [their] competitive side. I just want to beat the other people I see on there, so [using the app] is quite a good motivator.” [37]
- Quote 49: “It made me want to exercise more just, as like, kinda like, a competition to see how many calories because it takes your calories off whenever you exercise so I’m like let’s see how many I can get off this time.” [48]
- Quote 50: “Someone whose successful and quit smoking isn’t any better than someone that’s struggling with it. Like, no, I didn’t-I don’t like that aspect...it just makes someone feel bad.” [67]

Impersonated app

- Quote 51: “It’s like a ‘little boss in my pocket’... that’s sort of saying ‘you know you need to get out and do this’.” [45]
- Quote 52: “It’s like your own little motivator, in a way. And it definitely, it’s like, okay it’s like a little person, but it doesn’t talk, but it’s like, you shouldn’t eat that, or it’s like you should. So I don’t know it’s, I like it—I mean, I think it’s cool. It’s like my own little motivation.” [48]
- Quote 53: “I don’t want an electronic device telling me what to do.” [56]

Textbox 7. Illustrative quotes (Q54-Q59) for factors mapped onto automatic motivation subcomponents of the capability, opportunity, motivation, behavior model and coded under the theoretical domains framework: reinforcement and emotions.

Engagement

Feedback

- Quote 54: "I liked how it gave notifications, like every day I've got a notification saying; You're on day four of your smoking quitting history. You could do this, don't give up. Stay loyal and stuff like that. That was quite impressive." [72]
- Quote 55: "The big green continue at the bottom and when it moves on to the next thing I feel great, I've achieved something, I've filled something in correctly. I like that. And a nice little noise which made me think, Oh, I'm not an idiot." [46]
- Quote 56: "The progress I didn't make—it shows [and thus is demotivating]." [56]

Rewards

- Quote 57: "Earning badges [was] important when I was doing it...We learned as a kid, to consider [it] as [an] accomplishment." [56]
- Quote 58: "Each time you try, you get the points. And if these points can be converted to something else. Because you know, you're not really working for the badge but if the virtual badge can turn into something tangible, I would want that." [57]
- Quote 59: "Well, both of them are a kind of 'well done for doing this', they're both a reward, they both make you feel a bit better. But a badge, it's a cool fact, but it's not the same as having vouchers, where you can go and treat yourself to something you want." [59]

Theoretical Domains Framework Domain: Emotions

Emotions, based on previous experiences and behavior, are a complex reaction by which people tend to respond to a personally important event or matter (Textbox 8). Curiosity [38,52,54,61] positively influences the uptake of health and well-being smartphone apps (Q60). However, in 2 studies, both targeting alcohol consumption reduction, this factor was only relevant for a specific user type: for those who were characterized as *low-risk* drinkers [38] and *noncommitters* (ie, users who did not commit to engage with the app and, thus, did not gain any benefit from it) of the app [54].

Textbox 8. Illustrative quote (Q60) for factors mapped onto the automatic motivation subcomponent of the capability, opportunity, motivation, behavior model and coded under the theoretical domains framework: emotion.

Uptake

Curiosity

- Quote 60: "It was more like seeing an ad and just, okay I should try this — and then I found it on the internet and signed up. It was more like a fun thing. We'll see if it works. More like that." [52]

Reflective Motivation

Theoretical Domains Framework: Goals

Goals are outcomes that an individual would like to achieve to change a certain behavior (Textbox 9). Goal setting [38,39,45,48,51,54,56,58,59,66,71,74] was related to sustained engagement with health and well-being apps (Q61). Some users chose to set a goal, and mostly, this was only 1 goal at a time, so their focus would remain on 1 single aspect of change of the behavior (Q62), whereas others were more reluctant to use this feature because of fears of not being able to achieve their set goal and to avoid disappointing themselves (Q63) [38]. In general, the studies suggest that users were more determined to engage in behavior change when they had set goals [45] and believed they had successfully achieved or could achieve their goals with the help of an app by increasing their intention to use the app and by better monitoring the target behavior (Q64-Q65) [48,54,56,58,59].

Textbox 9. Illustrative quotes (Q61-Q65) for factors mapped onto the reflective motivation subcomponent of the capability, opportunity, motivation, behavior model and coded under the theoretical domains framework: goals.

Engagement

Goal setting

- Quote 61: "I'm not good at self-discipline and exercise, so maybe this [goal setting in the app] can help me get to my goal." [56]
- Quote 62: "I only set one goal because I was very keen to kind of remain focused on one thing. I didn't want to come and get lost in the app using it like a game. You know, I wanted to use it for one very specific thing... I think I set it to drink probably within guidelines." [38]
- Quote 63: "No, it didn't appeal - probably because I thought if I put some goals in I'm probably not going to stick to it, which probably makes me sound a bit naughty." [38]
- Quote 64: "If you set those manageable goals, so you could achieve it, if you feel like you're actually progressing, getting something, then you're more likely to go back." [58]
- Quote 65: "It would encourage me to open the app on a daily basis." [59]

Theoretical Domains Framework Domain: Beliefs About Consequences

This domain includes aspects related to outcome expectancies (Textbox 10). Perceived utility of the app [37,46,52,59,61,74] refers to where there is a discrepancy between what the users are looking for and what an app actually offers. It was suggested that the unmet expectations of an app would lead to disengagement and frustration with the app (Q66-Q68).

Textbox 10. Illustrative quotes (Q66-Q68) for factors mapped onto the reflective motivation subcomponent of the capability, opportunity, motivation, behavior model and coded under the theoretical domains framework: beliefs about consequences.

Engagement

Perceived utility of the app

- Quote 66: "I do have some apps I don't use often, mainly because they've kind of bored me in a way. I'll just do an example: one fitness app shows you how to lose weight, but the way it's describing it, it's not what I'm after. It's one of those free apps I bought that—I thought [the fitness app] would be great, but when you actually use it, it's not the same." [37]
- Quote 67: "I think that's where it let itself down for me. Once I'd played with it, once I tried the game, done the identity and whatnot, there wasn't much else there for me." [46]
- Quote 68: "It [mindfulness app] didn't add anything...I guess it didn't detract, it didn't make anything worse, but it didn't add anything to my armoury, I guess, my tool kit, as keeping myself sane, I suppose, it didn't add." [61]

Other Factors

There were a number of sociodemographic factors that did not fit clearly under the components of the COM-B model.

Sociodemographic Factors

Apps were more frequently downloaded by women than men, with the percentage ranging from 59% to 74% [38,41,49,53,55,63], although 1 study found that being male was associated with using an app to manage alcohol consumption [65]. Being younger than 44 years was associated with a higher level of uptake and engagement [38,41,42,44,49,53,55,63,64] than older adults. Living in an urban area [42,44,55]; having a better education level, such as having high school education or higher [41,42,44,64] and college degree or higher [41,53]; and having a higher income [44] were also associated with better engagement with health and well-being apps.

Discussion

Principal Findings

This is the first systematic review to conduct a theoretical analysis using the COM-B model of factors influencing the uptake of and engagement with health and well-being apps. The findings from this review suggest that there are 26 key factors across the constructs of capability, opportunity, and motivation that influence the uptake of and engagement with these types of apps, which were found to be important for a wide range of populations and behaviors.

Our review replicates previous findings in the wider literature on digital behavior change interventions. The core findings of our review suggest that attention should be perhaps shifted mainly to the support and guidance offered to new and existing users of health and well-being apps. We found that support and guidance of uptake can be targeted by increasing their awareness of health apps through, for example, recommendations received from health practitioners. In line with the findings of previous reviews, help with initial engagement could be achieved by improving the users' app literacy skills and by providing knowledge [14,17]. We present knowledge in a novel way by breaking it down to instructions on how to use it (ie, user guidance), advice related to the target behavior or condition (ie, health information), and information on their progress or data (ie, statistical information). This suggests that allowing access to users to different information that serves different purposes (eg, health benefits vs progress data) would enhance their engagement through different channels, such as guidance, support, and education.

Potentially, one of the most important factors for engagement identified in this review is health practitioner support. In line with the emerging evidence from the human-computer interaction (HCI) literature, we found that an app coupled with human support [14,17] was likely to be more effective by increasing the intervention effectiveness and engagement [78,79]. Alternatively, human support can be impersonated by embedded artificial intelligence (AI) features. A recent experimental study found that a supportive AI-powered chatbot doubled the engagement with a smoking cessation app and increased its effectiveness [80]. This suggests that embedded human support or features that mimic human support might lead to greater engagement with digital behavior change tools.

Behavior change techniques, widely reported by others previously [14,17-19], were also identified as important factors to sustain engagement, including self-monitoring, feedback, goal setting, reminders, rewards, and social support. However, we found that not all of these have a positive effect. Reminders and social support factors (embedded social media and social competition) are not universally useful and might cause disengagement or even harm by triggering negative emotions. One plausible explanation is that the participants of the studies included may or may not have real-life experience with health and well-being apps. Some of the included studies examined participants' perceptions about a hypothetical app or an app that was planned to be developed. These studies relied on the participants' opinion of what they think would be important for them in terms of uptake of and engagement with health and well-being apps, rather than sharing their lived experiences with such tools. For example, reminders were found useful in all the

studies targeting a hypothetical app, as opposed to those that were researching engagement with an app that had been used by the participants, where opinions about reminders were mixed, with some users finding them annoying. Another explanation is that the importance of these factors might be dependent on the target behavior. For example, people using apps that target mental health might not want to engage with social competition features or to share their progress or experiences on social media. This suggests that some of the identified factors in this review might be behavior dependent.

Another interesting finding, not identified in previous literature, is the safety netting characteristic of an app. This characteristic could promote long-term engagement rather than short goal-oriented engagement. The user could disengage at any time and reengage at a later stage when needed. This feature might be particularly useful for addiction research targeting relapse prevention strategies.

No factors were coded directly under 4 out of the 14 TDF domains (optimism, social identity, beliefs about capabilities, and intentions). However, 2 of these were highlighted in this review. We described how several factors coded under different domains affect intentions (eg, having adequate app literacy skills or user guidance provided to the user), in a manner similar to how emotions, other than curiosity, affect engagement with an app (eg, lack of app literacy skills triggers negative emotions, some found reminders annoying, or some fear of social comparison related to sharing on social media). We also found that aspects of the factor *personalization to needs* also include social identity aspects. Some communities (LGBTQ+ and cancer patients) prefer an app that is personalized to their social identity. Although social identity, in this case, was judged to be a weak factor to list it independently. In terms of the other two absent domains, factors under beliefs in their capabilities and optimism might be less relevant for uptake and engagement with health apps, or the studies may have missed them out, or, potentially, we failed to identify them from the included studies.

The importance of promoting equality and embracing cultural diversity has been partially identified previously [18]. Several studies in this review reported that apps should be provided at a low cost to users. It was suggested that multiculturalism should be embraced, and regional languages should be added. The concern of inequality for those who do not own a smartphone was also raised in this review [40]. An accompanying website was suggested as an alternative for homeless people who would not have access to a smartphone but may have access to the internet through nonprofit organizations, charities, or community libraries.

Strengths and Limitations

One major strength of this paper is that it adhered to the best practice processes for undertaking reviews by following the PRISMA guidance and Cochrane handbook [27,29]. By including all study designs, we were able to pool together and triangulate evidence and provide a novel and powerful synthesis of different study designs.

The use of theoretical frameworks is another strength. Other theoretical models were considered for this review, including the technology acceptance model [81] and the HCI models and theories [82]. However, the COM-B and TDF present advantages owing to their dynamic nature and by explaining the influences between components as they were developed from, and to represent, all theoretical components in behavior change–related models and theories. COM-B was explicitly developed to inform behavior change interventions through its connection to the Behavior Change Wheel [83], a tool that provides guidance on designing behavior change interventions. The factors identified under the components of the COM-B model allow easy identification of the intervention functions to target increased uptake of and engagement with health and well-being smartphone apps.

This review has several limitations. The review focused on 4 major behaviors related to prevention (smoking, alcohol consumption, physical activity, and diet) and mental health and well-being and could not capture other prevention type behaviors (eg, fall prevention). Factors relating to the uptake and engagement of apps focusing on other behaviors or conditions may differ from those found in this review and warrant further investigation.

Although we captured a wide range of populations, most of the included studies were carried out in high-income countries. Therefore, the findings might not be transferable to low- and middle-income countries or to other cultures. The quality of the studies was mixed. In some qualitative studies, the authors provided interpretations of their findings without an explicit quotation to support them. These interpretations were handled with care and were often ignored when no further explanation was provided about a concept. This might have led to losing some potentially important factors, not identified otherwise.

Policy and Practice: Recommendations and Implications

The findings of this review can inform app developers and researchers on how to develop health and well-being smartphone apps to better support behavior change and manage and monitor different physical and mental health conditions in adults.

This review may also have implications for policies that target prevention using digital technologies. Apps are an easy way to provide health-promoting behaviors and may play an important role in prevention strategies. For example, the UK government has recently published a Green Paper entitled *Advancing our health: prevention in the 2020s*, which shifted their focus from *cure to prevention*, committing to encourage the population to live a healthier life [84]. Additionally, the *Long Term Plan* policy document of the NHS in the United Kingdom dedicates an entire chapter to prevention programs and includes plans on digitally delivered methods to improve access to information, education, and intervention [85].

As part of prevention and health management strategies, the NHS and partners have created a pool of health and well-being apps for the individuals to access (NHS Apps Library). This research could help people access effective apps that people will remain engaged with, although the extent to which the population is open to use these portals for uptake is yet unknown and something worth investigating in the future.

A number of important themes are described in the projects and policy documents mentioned above. Some relate to digital health, for example, with an aim to reduce health inequalities [84] or to improve population health with personalized content and tailored lifestyle advice [85]. Our review suggests that app literacy skills are important for uptake. Enhancing app literacy skills for the elderly (eg, drop-in sessions in community settings) might be a feasible way to reduce health inequalities. Furthermore, some of the engagement-related factors might suggest the use of tailored lifestyle advice to address health behaviors, for example, by receiving personalized content within the app and web-based or offline help or advice from health practitioners as well as receiving recommendations for use of health apps from their health care professionals and general practitioner practices.

Therefore, our findings could inform stakeholders in public health, policy makers, and providers of health and well-being smartphone app portals to provide additional support for the uptake of and engagement with these digital interventions for adults.

Recommendations for stakeholders in public health, policy makers, and health and well-being app developers derived from the findings of this review can be found in Table 3.

Table 3. Recommendations for stakeholders in public health, policy, industry, health care, and health and well-being app development.

Component	Policy makers/industry/health care providers might want to consider	App developers might want to consider
Capability	• Improving app literacy skills • Increasing awareness of effective health and well-being apps, by advertising offline (eg, general practitioner practices) and web-based (eg, social media)	• Promoting less cognitive load by enabling automatization of data collection • Including user guidance that can be deactivated once the functionality of the app has been achieved (eg, help button) • Including content that targets education, health prevention, and health consequences related to the behavior that is targeted to change • Including statistical information (eg, graphs, percentages, and numbers) about the user's progress • Including well-designed reminders where the user can choose the time and frequency of receiving it • Including the self-monitoring feature that enables users to create routines • Including a *safety netting* feature that allows users to fall back on, even when the target behavior has been achieved
Opportunity	• Providing web-based or offline health practitioner support • Providing recommendations for health and well-being apps by health care professionals • Offering apps for free or at a low cost	• Allowing the provision of health professional support within the app • Allowing community networking within the app with other users • Organizing competition and challenges for users to opt in to • Avoiding automatic synching with the embedded social media (when applicable) • Personification of the app, by designing human-type attributes • Offering apps for free or at a low cost • Offering personalization of the app according to their demographics and individual and cultural needs
Motivation	• Offering tangible rewards, such as points that could be used as a discount in pharmacies or at other health- and well-being–related domains or health insurance providers • Providing a meaningful title and clear description of what the app does and what can offer, and how can help the user	• Providing positive, nonjudgmental, constructive, and informative feedback • Include gamification elements and offering rewards • Including goal-setting features (when applicable) • Providing a meaningful title and clear description of what the app does and what can offer, and how can help the user

Future Research

Although some of the factors identified and presented in the Results section appear to have a positive influence on uptake and engagement, there are mixed findings that might benefit from further investigation, such as reminders, embedded social media, and social competition. In the studies included in the review, descriptions of notification-type messages, such as reminders, feedback, push notifications, and other notifications, were used interchangeably, and it was not always clear which notifications were being referred to. Consistent terminology would help eliminate doubt around these concepts in the future. Issues around equality and diversity were highlighted in a few studies as something future research should address. Further

work is also needed to aid our understanding of how to avoid digital health widening inequalities through the exclusion of individuals who face a financial barrier to owning a smartphone or to purchasing an app, or who do not possess the skills to use one.

Conclusions

This is the first systematic review to investigate factors that influence the uptake of and engagement with health and well-being smartphone apps. We identified 26 factors that are relevant to a wide range of populations and different behaviors. These have clear implications for improving population health and targeting health inequalities. We provide a list of recommendations built on the identified factors to guide app developers, health app portal developers, and policy makers when commissioning, developing, and optimizing health and well-being smartphone apps. These can help address the issues of suboptimal uptake and engagement, which currently constrain the public health benefit of apps.

Acknowledgments

The authors would like to thank Mathew Smith, Academic Librarian for the Faculty of Medicine and Health Sciences at the University of East Anglia, and Nicola Pearce-Smith, senior information scientist from PHE, for their thoughts on the systematic search strategy. Furthermore, the authors are grateful to Dr Katherine Deane and to the University College London Tobacco and Alcohol Research Group for their expert opinion on this research. DS was funded through a Doctor of Philosophy studentship, provided jointly by PHE and the University of East Anglia.

Authors' Contributions

The initial concept was developed by DS, TC, and FN. DS wrote the study protocol with contributions from FN, AJ, TC, and JB. DS undertook data collection (literature search, screening, data extraction, and quality appraisal), data analysis, interpretation, and report writing. FN double checked the study selection, data extraction, and data coding. FN and AJ double assessed the quality of the included studies. DS prepared the paper. All authors read, commented, and contributed to the final paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
A visual representation of mapping the capability, opportunity, motivation, behavior model onto the Theoretical Domains Framework.
[PDF File (Adobe PDF File), 82 KB - jmir_v22i5e17572_app1.pdf]

Multimedia Appendix 2
Preferred Reporting Items for Systematic Reviews and Meta-Analyses checklist.
[PDF File (Adobe PDF File), 95 KB - jmir_v22i5e17572_app2.pdf]

Multimedia Appendix 3
Medical Literature Analysis and Retrieval System Online, or MEDLARS Online search strategy.
[PDF File (Adobe PDF File), 29 KB - jmir_v22i5e17572_app3.pdf]

Multimedia Appendix 4
Characteristics of the studies included in the review.
[PDF File (Adobe PDF File), 179 KB - jmir_v22i5e17572_app4.pdf]

Multimedia Appendix 5
Quality assessment of the studies included in the review.
[PDF File (Adobe PDF File), 168 KB - jmir_v22i5e17572_app5.pdf]

References

1. Yang Q, van Stee SK. The comparative effectiveness of mobile phone interventions in improving health outcomes: meta-analytic review. JMIR Mhealth Uhealth 2019 Apr 3;7(4):e11244 [FREE Full text] [doi: 10.2196/11244] [Medline: 30942695]
2. Schueller SM, Muñoz RF, Mohr DC. Realizing the potential of behavioral intervention technologies. Curr Dir Psychol Sci 2013 Dec 3;22(6):478-483. [doi: 10.1177/0963721413495872]

3. Yang G, Long J, Luo D, Xiao S, Kaminga AC. The characteristics and quality of mobile phone apps targeted at men who have sex with men in China: a window of opportunity for health information dissemination? JMIR Mhealth Uhealth 2019 Mar 27;7(3):e12573 [FREE Full text] [doi: 10.2196/12573] [Medline: 30916658]
4. Hou C, Carter B, Hewitt J, Francisa T, Mayor S. Do mobile phone applications improve glycemic control (HbA1c) in the self-management of diabetes? A systematic review, meta-analysis, and grade of 14 randomized trials. Diabetes Care 2016 Dec;39(11):2089-2095. [doi: 10.2337/dc16-0346] [Medline: 27926892]
5. Coorey GM, Neubeck L, Mulley J, Redfern J. Effectiveness, acceptability and usefulness of mobile applications for cardiovascular disease self-management: systematic review with meta-synthesis of quantitative and qualitative data. Eur J Prev Cardiol 2018 Mar;25(5):505-521. [doi: 10.1177/2047487317750913] [Medline: 29313363]
6. Schippers M, Adam PC, Smolenski DJ, Wong HT, de Wit JB. A meta-analysis of overall effects of weight loss interventions delivered via mobile phones and effect size differences according to delivery mode, personal contact, and intervention intensity and duration. Obes Rev 2017 Apr;18(4):450-459. [doi: 10.1111/obr.12492] [Medline: 28187246]
7. Semper HM, Povey R, Clark-Carter D. A systematic review of the effectiveness of smartphone applications that encourage dietary self-regulatory strategies for weight loss in overweight and obese adults. Obes Rev 2016 Sep;17(9):895-906. [doi: 10.1111/obr.12428] [Medline: 27192162]
8. Bardus M, van Beurden SB, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. Int J Behav Nutr Phys Act 2016 Mar 10;13:35 [FREE Full text] [doi: 10.1186/s12966-016-0359-9] [Medline: 26964880]
9. Meredith SE, Alessi SM, Petry NM. Smartphone applications to reduce alcohol consumption and help patients with alcohol use disorder: a state-of-the-art review. Adv Health Care Technol 2015;1:47-54 [FREE Full text] [doi: 10.2147/AHCT.S65791] [Medline: 27478863]
10. Song T, Qian S, Yu P. Mobile health interventions for self-control of unhealthy alcohol use: systematic review. JMIR Mhealth Uhealth 2019 Jan 29;7(1):e10899 [FREE Full text] [doi: 10.2196/10899] [Medline: 30694200]
11. Beyer F, Lynch E, Kaner E. Brief interventions in primary care: an evidence overview of practitioner and digital intervention programmes. Curr Addict Rep 2018;5(2):265-273 [FREE Full text] [doi: 10.1007/s40429-018-0198-7] [Medline: 29963364]
12. Rathbone AL, Prescott J. The use of mobile apps and SMS messaging as physical and mental health interventions: systematic review. J Med Internet Res 2017 Aug 24;19(8):e295 [FREE Full text] [doi: 10.2196/jmir.7740] [Medline: 28838887]
13. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: a systematic review. J Med Internet Res 2016 May 16;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
14. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
15. Kohl LF, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. J Med Internet Res 2013 Jul 16;15(7):e146 [FREE Full text] [doi: 10.2196/jmir.2665] [Medline: 23859884]
16. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
17. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? An evidence review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
18. Coughlin SS, Whitehead M, Sheats JQ, Mastromonico J, Smith S. A review of smartphone applications for promoting physical activity. Jacobs J Community Med 2016;2(1):21 [FREE Full text] [Medline: 27034992]
19. 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 Oct;131:70-81. [doi: 10.1016/j.diabres.2017.06.016] [Medline: 28692830]
20. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 May 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
21. Fulton EA, Brown KE, Kwah KL, Wild S. StopApp: using the behaviour change wheel to develop an app to increase uptake and attendance at NHS stop smoking services. Healthcare (Basel) 2016 Jul 8;4(2):E31 [FREE Full text] [doi: 10.3390/healthcare4020031] [Medline: 27417619]
22. Handley MA, Harleman E, Gonzalez-Mendez E, Stotland NE, Althavale P, Fisher L, et al. Applying the COM-B model to creation of an IT-enabled health coaching and resource linkage program for low-income Latina moms with recent gestational diabetes: the STAR MAMA program. Implement Sci 2016 May 18;11(1):73 [FREE Full text] [doi: 10.1186/s13012-016-0426-2] [Medline: 27193580]
23. Tombor I, Shahab L, Brown J, Crane D, Michie S, West R. Development of SmokeFree baby: a smoking cessation smartphone app for pregnant smokers. Transl Behav Med 2016 Dec;6(4):533-545 [FREE Full text] [doi: 10.1007/s13142-016-0438-0] [Medline: 27699682]

24. Atkins L, Francis J, Islam R, O'Connor D, Patey A, Ivers N, et al. A guide to using the theoretical domains framework of behaviour change to investigate implementation problems. Implement Sci 2017 Jun 21;12(1):77 [FREE Full text] [doi: 10.1186/s13012-017-0605-9] [Medline: 28637486]
25. Craig LE, McInnes E, Taylor N, Grimley R, Cadilhac DA, Considine J, et al. Identifying the barriers and enablers for a triage, treatment, and transfer clinical intervention to manage acute stroke patients in the emergency department: a systematic review using the theoretical domains framework (TDF). Implement Sci 2016 Nov 28;11(1):157 [FREE Full text] [doi: 10.1186/s13012-016-0524-1] [Medline: 27894313]
26. Heslehurst N, Newham J, Maniatopoulos G, Fleetwood C, Robalino S, Rankin J. Implementation of pregnancy weight management and obesity guidelines: a meta-synthesis of healthcare professionals' barriers and facilitators using the theoretical domains framework. Obes Rev 2014 Jul;15(6):462-486. [doi: 10.1111/obr.12160] [Medline: 24629076]
27. 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):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
28. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. Fam Med 2005 May;37(5):360-363 [FREE Full text] [Medline: 15883903]
29. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Hoboken, New Jersey, United States: Wiley; 2008.
30. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. The mixed methods appraisal tool (MMAT) version 2018 for information professionals and researchers. Educ Inform 2018 Dec 18;34(4):285-291. [doi: 10.3233/efi-180221]
31. Higgins JT, Sterne JA, Savovic J, Page MJ, Hróbjartsson A, Boutron I, et al. A revised tool for assessing risk of bias in randomized trials. Cochrane Database Syst Rev 2016;10(Suppl 1):29-31 [FREE Full text]
32. Pluye P, Gagnon M, Griffiths F, Johnson-Lafleur J. A scoring system for appraising mixed methods research, and concomitantly appraising qualitative, quantitative and mixed methods primary studies in mixed studies reviews. Int J Nurs Stud 2009 May;46(4):529-546. [doi: 10.1016/j.ijnurstu.2009.01.009] [Medline: 19233357]
33. Sandelowski M. What's in a name? Qualitative description revisited. Res Nurs Health 2010 Mar;33(1):77-84. [doi: 10.1002/nur.20362] [Medline: 20014004]
34. Dixon-Woods M, Agarwal S, Jones D, Young B, Sutton A. Synthesising qualitative and quantitative evidence: a review of possible methods. J Health Serv Res Policy 2005 Jan;10(1):45-53. [doi: 10.1177/135581960501000110] [Medline: 15667704]
35. Sandelowski M, Voils CI, Barroso J. Defining and designing mixed research synthesis studies. Res Sch 2006;13(1):29 [FREE Full text] [Medline: 20098638]
36. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008 Jul 10;8:45 [FREE Full text] [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]
37. Anderson K, Burford O, Emmerton L. Mobile health apps to facilitate self-care: a qualitative study of user experiences. PLoS One 2016;11(5):e0156164 [FREE Full text] [doi: 10.1371/journal.pone.0156164] [Medline: 27214203]
38. Attwood S, Parke H, Larsen J, Morton KL. Using a mobile health application to reduce alcohol consumption: a mixed-methods evaluation of the drinkaware track & calculate units application. BMC Public Health 2017 May 17;17(1):394 [FREE Full text] [doi: 10.1186/s12889-017-4358-9] [Medline: 28511698]
39. Baretta D, Perski O, Steca P. Exploring users' experiences of the uptake and adoption of physical activity apps: longitudinal qualitative study. JMIR Mhealth Uhealth 2019 Feb 8;7(2):e11636 [FREE Full text] [doi: 10.2196/11636] [Medline: 30735143]
40. Baskerville NB, Dash D, Wong K, Shuh A, Abramowicz A. Perceptions toward a smoking cessation app targeting LGBTQ+ youth and young adults: a qualitative framework analysis of focus groups. JMIR Public Health Surveill 2016 Dec 18;2(2):e165 [FREE Full text] [doi: 10.2196/publichealth.6188] [Medline: 27864164]
41. Bender MS, Choi J, Arai S, Paul SM, Gonzalez P, Fukuoka Y. Digital technology ownership, usage, and factors predicting downloading health apps among Caucasian, Filipino, Korean, and Latino Americans: the digital link to health survey. JMIR Mhealth Uhealth 2014 Oct 22;2(4):e43 [FREE Full text] [doi: 10.2196/mhealth.3710] [Medline: 25339246]
42. Bhuyan SS, Lu N, Chandak A, Kim H, Wyant D, Bhatt J, et al. Use of mobile health applications for health-seeking behavior among US adults. J Med Syst 2016 Jul;40(6):153. [doi: 10.1007/s10916-016-0492-7] [Medline: 27147516]
43. Bidargaddi N, Almirall D, Murphy S, Nahum-Shani I, Kovalcik M, Pituch T, et al. To prompt or not to prompt? A microrandomized trial of time-varying push notifications to increase proximal engagement with a mobile health app. JMIR Mhealth Uhealth 2018 Dec 29;6(11):e10123 [FREE Full text] [doi: 10.2196/10123] [Medline: 30497999]
44. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J Med Internet Res 2017 Apr 19;19(4):e125 [FREE Full text] [doi: 10.2196/jmir.5604] [Medline: 28428170]
45. Casey M, Hayes PS, Glynn F, OLaighin G, Heaney D, Murphy AW, et al. Patients' experiences of using a smartphone application to increase physical activity: the SMART MOVE qualitative study in primary care. Br J Gen Pract 2014 Aug;64(625):e500-e508 [FREE Full text] [doi: 10.3399/bjgp14X680989] [Medline: 25071063]

46. Crane D, Garnett C, Brown J, West R, Michie S. Factors influencing usability of a smartphone app to reduce excessive alcohol consumption: think aloud and interview studies. Front Public Health 2017;5:39 [FREE Full text] [doi: 10.3389/fpubh.2017.00039] [Medline: 28421175]
47. Gorton D, Dixon R, Maddison R, Mhurchu CN, Jull A. Consumer views on the potential use of mobile phones for the delivery of weight-loss interventions. J Hum Nutr Diet 2011 Dec;24(6):616-619. [doi: 10.1111/j.1365-277X.2011.01163.x] [Medline: 21781188]
48. Gowin M, Cheney M, Gwin S, Wann TF. Health and fitness app use in college students: a qualitative study. Am J Health Educ 2015 Jul 6;46(4):223-230. [doi: 10.1080/19325037.2015.1044140]
49. Guertler D, Vandelanotte C, Kirwan M, Duncan MJ. Engagement and nonusage attrition with a free physical activity promotion program: the case of 10,000 steps Australia. J Med Internet Res 2015 Jul 15;17(7):e176 [FREE Full text] [doi: 10.2196/jmir.4339] [Medline: 26180040]
50. Laurie J, Blandford A. Making time for mindfulness. Int J Med Inform 2016 Dec;96:38-50. [doi: 10.1016/j.ijmedinf.2016.02.010] [Medline: 26965526]
51. Lieffers JR, Arocha JF, Grindrod K, Hanning RM. Experiences and perceptions of adults accessing publicly available nutrition behavior-change mobile apps for weight management. J Acad Nutr Diet 2018 Feb;118(2):229-39.e3. [doi: 10.1016/j.jand.2017.04.015] [Medline: 28625662]
52. Ly KH, Janni E, Wrede R, Sedem M, Donker T, Carlbring P, et al. Experiences of a guided smartphone-based behavioral activation therapy for depression: a qualitative study. Internet Interv 2015 Mar;2(1):60-68. [doi: 10.1016/j.invent.2014.12.002]
53. Mackert M, Mabry-Flynn A, Champlin S, Donovan EE, Pounders K. Health literacy and health information technology adoption: the potential for a new digital divide. J Med Internet Res 2016 Oct 4;18(10):e264 [FREE Full text] [doi: 10.2196/jmir.6349] [Medline: 27702738]
54. Milward J, Deluca P, Drummond C, Kimergård A. Developing typologies of user engagement with the BRANCH alcohol-harm reduction smartphone app: qualitative study. JMIR Mhealth Uhealth 2018 Dec 13;6(12):e11692 [FREE Full text] [doi: 10.2196/11692] [Medline: 30545806]
55. Mitchell M, White L, Oh P, Alter D, Leahey T, Kwan M, et al. Uptake of an incentive-based mhealth app: process evaluation of the carrot rewards app. JMIR Mhealth Uhealth 2017 May 30;5(5):e70 [FREE Full text] [doi: 10.2196/mhealth.7323] [Medline: 28559224]
56. Peng W, Kanthawala S, Yuan S, Hussain SA. 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]
57. Peng W, Yuan S, Holtz BE. Exploring the challenges and opportunities of health mobile apps for individuals with type 2 diabetes living in rural communities. Telemed J E Health 2016 Sep;22(9):733-738. [doi: 10.1089/tmj.2015.0180] [Medline: 26982017]
58. Perski O, Blandford A, Ubhi HK, West R, Michie S. Smokers' and drinkers' choice of smartphone applications and expectations of engagement: a think aloud and interview study. BMC Med Inform Decis Mak 2017 Feb 28;17(1):25 [FREE Full text] [doi: 10.1186/s12911-017-0422-8] [Medline: 28241759]
59. Perski O, Baretta D, Blandford A, West R, Michie S. Engagement features judged by excessive drinkers as most important to include in smartphone applications for alcohol reduction: a mixed-methods study. Digit Health 2018;4:2055207618785841 [FREE Full text] [doi: 10.1177/2055207618785841] [Medline: 31463077]
60. Peters D, Deady M, Glozier N, Harvey S, Calvo RA. Worker preferences for a mental health app within male-dominated industries: participatory study. JMIR Ment Health 2018 May 25;5(2):e30 [FREE Full text] [doi: 10.2196/mental.8999] [Medline: 29695371]
61. Pung A, Fletcher SL, Gunn JM. Mobile app use by primary care patients to manage their depressive symptoms: qualitative study. J Med Internet Res 2018 Sep 27;20(9):e10035 [FREE Full text] [doi: 10.2196/10035] [Medline: 30262449]
62. Puszkiewicz P, Roberts AL, Smith L, Wardle J, Fisher A. Assessment of cancer survivors' experiences of using a publicly available physical activity mobile application. JMIR Cancer 2016 May 31;2(1):e7 [FREE Full text] [doi: 10.2196/cancer.5380] [Medline: 28410168]
63. Serrano KJ, Coa KI, Yu M, Wolff-Hughes DL, Atienza AA. Characterizing user engagement with health app data: a data mining approach. Transl Behav Med 2017 Jun;7(2):277-285 [FREE Full text] [doi: 10.1007/s13142-017-0508-y] [Medline: 28616846]
64. Sharpe JD, Zhou Z, Escobar-Viera CG, Morano JP, Lucero RJ, Ibañez GE, et al. Interest in using mobile technology to help self-manage alcohol use among persons living with the human immunodeficiency virus: A Florida Cohort cross-sectional study. Subst Abus 2018 Jan 02;39(1):77-82 [FREE Full text] [doi: 10.1080/08897077.2017.1356793] [Medline: 28723300]
65. Smahel D, Elavsky S, Machackova H. Functions of mhealth applications: a user's perspective. Health Informatics J 2019 Sep;25(3):1065-1075. [doi: 10.1177/1460458217740725] [Medline: 29121831]
66. Solbrig L, Jones R, Kavanagh D, May J, Parkin T, Andrade J. People trying to lose weight dislike calorie counting apps and want motivational support to help them achieve their goals. Internet Interv 2017 Mar;7:23-31 [FREE Full text] [doi: 10.1016/j.invent.2016.12.003] [Medline: 28286739]

67. Struik LL, Bottorff JL, Baskerville NB, Oliffe JL. The crush the crave quit smoking app and young adult smokers: qualitative case study of affordances. JMIR Mhealth Uhealth 2018 Jul 8;6(6):e134 [FREE Full text] [doi: 10.2196/mhealth.9489] [Medline: 29884602]
68. Sun L, Wang Y, Greene B, Xiao Q, Jiao C, Ji M, et al. Facilitators and barriers to using physical activity smartphone apps among Chinese patients with chronic diseases. BMC Med Inform Decis Mak 2017 May 19;17(1):44 [FREE Full text] [doi: 10.1186/s12911-017-0446-0] [Medline: 28420355]
69. Switsers L, Dauwe A, Vanhoudt A, van Dyck H, Lombaerts K, Oldenburg J. Users' perspectives on mhealth self-management of bipolar disorder: qualitative focus group study. JMIR Mhealth Uhealth 2018 May 2;6(5):e108 [FREE Full text] [doi: 10.2196/mhealth.9529] [Medline: 29720363]
70. Taki S, Russell CG, Lymer S, Laws R, Campbell K, Appleton J, et al. A mixed methods study to explore the effects of program design elements and participant characteristics on parents' engagement with an mhealth program to promote healthy infant feeding: the growing healthy program. Front Endocrinol (Lausanne) 2019;10:397 [FREE Full text] [doi: 10.3389/fendo.2019.00397] [Medline: 31293515]
71. Tang J, Abraham C, Stamp E, Greaves C. How can weight-loss app designers' best engage and support users? A qualitative investigation. Br J Health Psychol 2015 Mar;20(1):151-171. [doi: 10.1111/bjhp.12114] [Medline: 25130682]
72. Tudor-Sfetea C, Rabee R, Najim M, Amin N, Chadha M, Jain M, et al. Evaluation of two mobile health apps in the context of smoking cessation: qualitative study of cognitive behavioral therapy (CBT) versus non-CBT-based digital solutions. JMIR Mhealth Uhealth 2018 May 18;6(4):e98 [FREE Full text] [doi: 10.2196/mhealth.9405] [Medline: 29669708]
73. Wang N, Deng Z, Wen LM, Ding Y, He G. Understanding the use of smartphone apps for health information among pregnant Chinese women: mixed methods study. JMIR Mhealth Uhealth 2019 Jun 18;7(6):e12631 [FREE Full text] [doi: 10.2196/12631] [Medline: 31215516]
74. Webcredible. Presentation of their findings on digital healthcare for Public Health England (PHE). Unpublished 2016.
75. Woldaregay AZ, Issom D, Henriksen A, Marttila H, Mikalsen M, Pfuhl G, et al. Motivational factors for user engagement with mhealth apps. Stud Health Technol Inform 2018;249:151-157. [doi: 10.3233/978-1-61499-868-6-151] [Medline: 29866972]
76. Xie Z, Nacioglu A, Or C. Prevalence, demographic correlates, and perceived impacts of mobile health app use amongst Chinese adults: cross-sectional survey study. JMIR Mhealth Uhealth 2018 Apr 26;6(4):e103 [FREE Full text] [doi: 10.2196/mhealth.9002] [Medline: 29699971]
77. Zeng EY, Vilardaga R, Heffner JL, Mull KE, Bricker JB. Predictors of utilization of a novel smoking cessation smartphone app. Telemed J E Health 2015 Dec;21(12):998-1004 [FREE Full text] [doi: 10.1089/tmj.2014.0232] [Medline: 26171733]
78. Tate DF, Jackvony EH, Wing RR. Effects of internet behavioral counseling on weight loss in adults at risk for type 2 diabetes: a randomized trial. J Am Med Assoc 2003 May 9;289(14):1833-1836. [doi: 10.1001/jama.289.14.1833] [Medline: 12684363]
79. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to ehealth interventions. J Med Internet Res 2011 Mar 10;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
80. Perski O, Crane D, Beard E, Brown J. Does the addition of a supportive chatbot promote user engagement with a smoking cessation app? An experimental study. Digit Health 2019;5:2055207619880676 [FREE Full text] [doi: 10.1177/2055207619880676] [Medline: 31620306]
81. Holden RJ, Karsh B. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Mar;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
82. Carroll JM. HCI Models, Theories, and Frameworks: Toward a Multidisciplinary Science. Burlington, Massachusetts: Morgan Kaufmann; 2003.
83. Michie S, Atkins L, West R. The Behaviour Change Wheel: A Guide to Designing Interventions. London, UK: Silverback Publishing; 2014.
84. The Government of UK. 2019 Jul 22. Advancing Our Health: Prevention in the 2020s–Consultation Document URL: https://www.gov.uk/government/consultations/advancing-our-health-prevention-in-the-2020s/advancing-our-health-prevention-in-the-2020s-consultation-document [accessed 2020-04-01]
85. National Health Service. NHS Long Term Plan. 2019. URL: https://www.longtermplan.nhs.uk/ [accessed 2020-04-01]

Abbreviations

COM-B model: capability, opportunity, motivation, behavior model
TDF: theoretical domains framework
HCI: human-computer Interaction
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
MMAT: mixed methods appraisal tool
LGBTQ+: lesbian, gay, bisexual, transgender, queer, and other spectrum of sexuality and gender
MEDLINE: Medical Literature Analysis and Retrieval System Online, or MEDLARS Online

NHS: National Health Service
PHE: Public Health England

Edited by G Eysenbach; submitted 20.12.19; peer-reviewed by P Ning, S Taki; comments to author 10.03.20; revised version received 23.03.20; accepted 23.03.20; published 29.05.20.

Please cite as:
Szinay D, Jones A, Chadborn T, Brown J, Naughton F
Influences on the Uptake of and Engagement With Health and Well-Being Smartphone Apps: Systematic Review
J Med Internet Res 2020;22(5):e17572
URL: http://www.jmir.org/2020/5/e17572/
doi:10.2196/17572
PMID:32348255

Original Paper

Cost-Effectiveness of Virtual Reality Cognitive Behavioral Therapy for Psychosis: Health-Economic Evaluation Within a Randomized Controlled Trial

Roos Pot-Kolder[1], MSc; Wim Veling[2], MD, PhD; Chris Geraets[2], MSc; Joran Lokkerbol[3,4], PhD; Filip Smit[1,3,5], PhD; Alyssa Jongeneel[1,6], MSc; Helga Ising[6], PhD; Mark van der Gaag[1,6], PhD

[1]Department of Clinical, Neuro and Developmental Psychology, Vrije Universiteit, Amsterdam, Netherlands

[2]Department of Psychiatry, University Medical Centre Groningen, University of Groningen, Groningen, Netherlands

[3]Centre of Economic Evaluation and Machine Learning, Trimbos Institute, Utrecht, Netherlands

[4]Department of Health Care Policy, Harvard Medical School, Boston, MA, United States

[5]Department of Epidemiology and Biostatistics, University Medical Centers Amsterdam, Amsterdam, Netherlands

[6]Department of Psychosis Research, Parnassia Psychiatric Institute, The Hague, Netherlands

Corresponding Author:
Roos Pot-Kolder, MSc
Department of Clinical, Neuro and Developmental Psychology
Vrije Universiteit
Van der Boechorststraat 7
Amsterdam, 1081 BT
Netherlands
Phone: 31 641034764
Email: roos.potkolder@gmail.com

Abstract

Background: Evidence was found for the effectiveness of virtual reality-based cognitive behavioral therapy (VR-CBT) for treating paranoia in psychosis, but health-economic evaluations are lacking.

Objective: This study aimed to determine the short-term cost-effectiveness of VR-CBT.

Methods: The health-economic evaluation was embedded in a randomized controlled trial evaluating VR-CBT in 116 patients with a psychotic disorder suffering from paranoid ideation. The control group (n=58) received treatment as usual (TAU) for psychotic disorders in accordance with the clinical guidelines. The experimental group (n=58) received TAU complemented with add-on VR-CBT to reduce paranoid ideation and social avoidance. Data were collected at baseline and at 3 and 6 months postbaseline. Treatment response was defined as a pre-post improvement of symptoms of at least 20% in social participation measures. Change in quality-adjusted life years (QALYs) was estimated by using Sanderson et al's conversion factor to map a change in the standardized mean difference of Green's Paranoid Thoughts Scale score on a corresponding change in utility. The incremental cost-effectiveness ratios were calculated using 5000 bootstraps of seemingly unrelated regression equations of costs and effects. The cost-effectiveness acceptability curves were graphed for the costs per treatment responder gained and per QALY gained.

Results: The average mean incremental costs for a treatment responder on social participation ranged between €8079 and €19,525, with 90.74%-99.74% showing improvement. The average incremental cost per QALY was €48,868 over the 6 months of follow-up, with 99.98% showing improved QALYs. Sensitivity analyses show costs to be lower when relevant baseline differences were included in the analysis. Average costs per treatment responder now ranged between €6800 and €16,597, while the average cost per QALY gained was €42,030.

Conclusions: This study demonstrates that offering VR-CBT to patients with paranoid delusions is an economically viable approach toward improving patients' health in a cost-effective manner. Long-term effects need further research.

Trial Registration: International Standard Randomised Controlled Trial Number (ISRCTN) 12929657; http://www.isrctn.com/ISRCTN12929657

(J Med Internet Res 2020;22(5):e17098) doi:10.2196/17098

KEYWORDS

psychosis; virtual reality; cognitive behavioral therapy; cost-effectiveness

Introduction

Psychotic disorders impose a large disease burden—morbidity plus mortality—on the population, and in its wake, substantial economic costs occur for society and health care systems. The main drivers of societal costs of schizophrenia are health care costs and productivity losses, but patients and their families also incur substantial costs [1]. Low participation rates of individuals with psychosis in the labor market are an important cause of productivity losses, while the main contributor to health care costs are in-patient psychiatric admissions [2]. All in all, treatment costs of psychotic disorders consume a significant part of health care budgets in European countries [3].

Paranoid ideation is a common delusion in individuals with a psychotic disorder. Even when medicinal treatment is successful, paranoid thoughts and anxiety often remain because of conditioned avoidance and other acquired safety behaviors in social situations [4]. Social avoidance hinders recovery in social participation for patients and keeps unemployment rates as high as 70%-85% [5,6]. A poor social network contributes to stigma and a lack of empowerment, resulting in more depressive symptoms and lower quality of life [7]. A smaller social network size is associated with more severe overall psychiatric and negative symptoms [8]. Virtual reality-based cognitive behavioral therapy (VR-CBT) was found to be an effective treatment for paranoid ideation in individuals with a psychotic disorder [9,10]. The use of virtual reality (VR) treatment in clinical practice is expected to become more widespread as VR technology becomes more readily available [11]. Therefore, information on the cost-effectiveness of this kind of treatment is required. This study was designed to evaluate whether adding VR-CBT to treatment as usual (TAU) would be effective in treating paranoid ideation in a cost-effective way with respect to improving social participation. A trial-based cost-effectiveness analysis (CEA) was conducted using data collected in seven outpatient mental health care services in the Netherlands, comparing add-on VR-CBT with TAU alone. This paper aims to determine the short-term (ie, 6-month) cost-effectiveness of VR-CBT from a societal perspective.

Methods

Research Design

The health-economic evaluation was embedded in a randomized controlled trial evaluating VR-CBT in 116 patients with a psychotic disorder suffering from paranoid ideation [10]. The VR-CBT study was a randomized, controlled, single-blind multicenter trial in two parallel groups, comparing add-on VR-CBT to TAU alone over a period of 6 months [10]. This study was approved by the Vrije Universiteit (VU University) Medical Ethics Committee for mental health service research and was registered retrospectively at the ISRCTN (International Standard Randomised Controlled Trial Number) registry (ISRCTN12929657). The trial protocol is provided elsewhere [12]. Four virtual social environments—a street, bus, café, and supermarket—were created with Vizard software (WorldViz). Within the environment, participants could move by operating a Logitech F310 Gamepad. They used a Sony HMZ-T1/T2/T3 head-mounted display with a high-definition resolution of 1280 × 720 per eye, a 51.6 diagonal field of view, and a 3DOF (3 degrees of freedom) tracker for head rotation. VR-CBT therapists were psychologists with at least basic cognitive behavioral therapy (CBT) training. They received 2 days of training in VR-CBT. The VR-CBT manual described a structured treatment plan for all 16 sessions. Therapists were supervised in a group for 4 hours every month by two VR-CBT specialists.

Recruitment

Participants were recruited at seven treatment centers in the Netherlands between April 1, 2014, and December 31, 2015. To be included, participants had to meet the following criteria: (1) 18-65 years of age; (2) DSM-IV (Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition) diagnosis of schizophrenia, schizophreniform disorder, schizoaffective disorder, delusional disorder, or psychotic disorder not otherwise specified; (3) suffering from at least mild paranoia, as assessed by Green's Paranoid Thoughts Scale (GPTS) (score of >40); and (4) self-report of avoiding at least one social situation. Exclusion criteria were as follows: (1) insufficient competency of Dutch language; (2) IQ below 70; and (3) a concurrent diagnosis of epilepsy. Assessments were performed at baseline and at 3 and 6 months postbaseline.

Interventions

All participants continued to receive TAU (ie, antipsychotic medication, regular contact with a psychiatrist to manage symptoms, and regular contact with a psychiatric nurse). Participants in the experimental condition also received therapist-led VR-CBT. VR-CBT treatment consisted of 16 biweekly sessions of 60 minutes each, using 40 minutes for exposure and behavioral exercises in virtual social environments. The therapist used an individual case formulation to help patients falsify their harm expectancies. No homework exercises were given between VR-CBT sessions. The treatment protocol, in Dutch, is available from the corresponding author.

Outcome Measures

Overview

We conducted both a CEA with three measures of improved social participation as outcome and a cost-utility analysis (CUA) with quality-adjusted life years (QALYs) gained as outcome. The outcome measures are described in more detail below.

Social Participation

The outcome of interest in the CEA was social participation. Social participation was operationalized in three ways: (1) objective social participation as the amount of time spent with others, (2) subjective social participation as momentary anxiety, and (3) subjective social participation as momentary paranoia. *Momentary* in this context meant that it was measured in real

life during social company. All three outcomes were assessed in real time using the ecological sampling method (ESM). ESM is a structured diary method in which individuals are asked in daily life to report their thoughts, feelings, and symptoms, as well as the appraisal of the present social context. To that end, all participants carried an electronic device (PsyMate) for the ESM assessments. The device beeped at semirandom moments 10 times a day over 6 days. At each beep, the device collected self-assessments on a 7-point Likert scale ranging from 1 (not at all) to 7 (very). A positive treatment response on each of the three outcome measures was defined as an improvement of at least 20% at 6 months follow-up relative to the patient's baseline score.

Quality-Adjusted Life Year

The outcome in the CUA was the QALY derived from the GPTS [13]. The GPTS is an established broad measure of paranoid-delusional functioning that has long been used as an outcome measure. This instrument was chosen to be able to compare results with earlier CUA research on the subject. Mean GPTS scores at each measurement were first converted into the standard mean difference (SMD) by dividing the raw mean change scores by the SD of the GPTS at baseline in the control condition. In a next step, we used Sanderson et al's conversion factor [14] of 0.1835 (ie, the average of 0.209 using a rating scale and 0.158 using time trade-off), such that a change of 1 standard unit (ie, SMD) on the GPTS is equal to a corresponding change of 0.18 utility. The utility is a quality of life valuation and is needed to compute QALY gains in the VR-CBT condition relative to the TAU condition over the full 6 months between baseline and follow-up.

Resource Use and Costing

Societal costs were computed by adding (1) the direct medical costs of health care services use including the costs of antipsychotic medication and, in the experimental condition, the additional costs of adjunctive VR-CBT treatment; (2) direct nonmedical costs of travel; and (3) indirect costs stemming from lower productivity. For each participant, cost data over the last 3 months were collected at each of three measurement points. Resource use data, for costing, were collected using the Trimbos Institute and Institute of Medical Technology Assessment Questionnaire for Costs Associated with Psychiatric Illness (TiC-P) [15]. The TiC-P is the most widely used health service interview in the Netherlands. It consists of questions on the number of contacts by type of health care provider and questions on productivity losses. A health service questionnaire is a valid and reliable method for quantifying costs in trial-based economic evaluations in health care [16]. A cross-validation sample comparing TiC-P self-report to electronic patient files showed all data to be reliable, except for the number of reported sessions with a psychologist (data available upon request from first author). Not all patients had understood that they needed to incorporate the 16 VR-CBT sessions into their TiC-P self-report. This information was, therefore, 100% cross-checked using electronic patient files. The main cost driver was admission to psychiatric hospitals, so the number of days admitted to a psychiatric hospital was also 100% cross-checked against electronic patient files and corrected where needed.

Direct medical costs were calculated by multiplying health service units (eg, sessions, visits, and hospital days) with their standard economic cost price (see Multimedia Appendix 1). We also added the medication costs, consisting of antipsychotic and antidepressant medication. Corresponding costs were calculated as the cost price per standard daily dose, as reported in the Dutch Pharmaceutical Compass [17], multiplied by the number of prescription days, plus the pharmacist's dispensing costs of €6 per monthly prescription or €12 for a first-time prescription [18].

Virtual Reality Costs

For VR therapy hardware, software and training costs were calculated. Total yearly costs for one VR system was €23,995, according to CleVR BV, a company who builds VR sets. Yearly costs for training and supervision of the psychologists was €13,400. Per-patient costs per 16 VR-CBT treatment sessions was €373.95.

Travel Costs

Travel costs arose when participants had to make return trips for receiving health care at health services. Travel costs were computed as the average distance to a health service (7 km) multiplied by the costs per km (€0.21) [18].

Productivity Costs

Research assistants monitored changes in the participants' work status at baseline and at 3 and 6 months postbaseline using the TiC-P. Productivity losses in paid work were calculated according to the human capital approach [19], reflecting changes in the contractual number of hours worked per week and adjusting these for work-loss days arising from sick leave over the full period of 6 months using gender-specific hourly productivity costs [18]. Costs were originally expressed in Euros for the reference year 2014, but indexed to 2015 using the consumer price index as reported by Statistics Netherlands. In the reference year 2015, 1 Euro in the Netherlands equaled 1.235 US$.

Statistical Analysis

Imputation

Following the CONSORT (Consolidated Standards of Reporting Trials) and CHEERS (Consolidated Health Economic Evaluation Reporting Standards) guidelines, all our analyses adhered to the intention-to-treat principle. To that end, missing values were imputed using multiply imputed chained equations (MICE) for nonparametric data with M of 100 bootstraps for each incomplete variable. Baseline variables predictive of effects (ie, QALYs and treatment response) were used for imputation, such as baseline data of the variable with missing values, treatment condition, ethnicity, education, sex, age, and safety behaviors at baseline. Safety behaviors, such as avoiding eye contact or escaping from social situations, were measured using the Safety Behaviour Questionnaire-Persecutory Delusions (SBQ-PD) [20]. *Time spent with others* showed a large difference at baseline despite randomization and was added as covariate in the CEA where *time spent with others* was used as the treatment response outcome of interest.

Main Analysis

Both the CUA and CEA were conducted from the societal perspective. In each of these analyses, the incremental cost-effectiveness ratios (ICERs) were calculated as the between-group cost difference divided by the between-group effect difference. The ICER is interpreted as the additional costs per additional unit effect (ie, per additional treatment responder; per QALY gained). Cost and effect differences were obtained from seemingly unrelated regression equations of costs and effects, thus allowing for correlated residuals in the equations. The seemingly unrelated regression equations (SURE) models were bootstrapped 5000 times. In each bootstrap step, the mean cost differences and the mean outcome differences were computed and these were plotted on the cost-effectiveness plane. Finally, cost-effectiveness acceptability curves (CEACs) were graphed. CEACs inform decision makers about the likelihood that an intervention is deemed cost-effective, given a range of willingness-to-pay ceilings for gaining 1 QALY and gaining 1 treatment responder. All analyses were conducted in Stata, version 13.1 (StataCorp).

Sensitivity Analyses

The following sensitivity analyses were carried out. First, a sensitivity analysis was done including safety behavior at baseline as a covariate because despite randomization there was a significant difference at baseline, and it was found to be the main mediator in reducing paranoid ideation [10]. Second, a sensitivity analysis was done including psychiatric admission costs at baseline as a covariate because there was a large difference between groups at baseline. Third, a sensitivity analysis was done including both safety behavior at baseline and psychiatric admission costs at baseline as covariates in the model.

Results

Overview

After providing informed consent, 116 participants agreed to participate: 58 (50.0%) in the control condition and 58 (50.0%) in the experimental condition (see Figure 1).

Baseline characteristics of the sample can be found in Table 1. Results of costs and outcomes can be found in Table 2. A small group of participants was responsible for a large portion of the baseline costs, largely related to hospital admissions. The total days of psychiatric admissions were 233 days at baseline, 101 days posttreatment, and zero days at follow-up for the VR-CBT group. The total days of psychiatric admissions were 138 days at baseline, 20 days posttreatment, and 68 days at follow-up for the TAU group.

Figure 1. Trial flow diagram. *Specification of participants lost to posttreatment: 6 declined further participation and 2 were lost due to clerical errors by therapist. ‡Specification of participants lost to follow-up: 9 declined further participation, 1 died of unrelated causes, and 2 were lost due to clerical errors by therapist. First published in Lancet Psychiatry (Pot-Kolder et al, 2018). VR-CBT: virtual reality-based cognitive behavioral therapy.

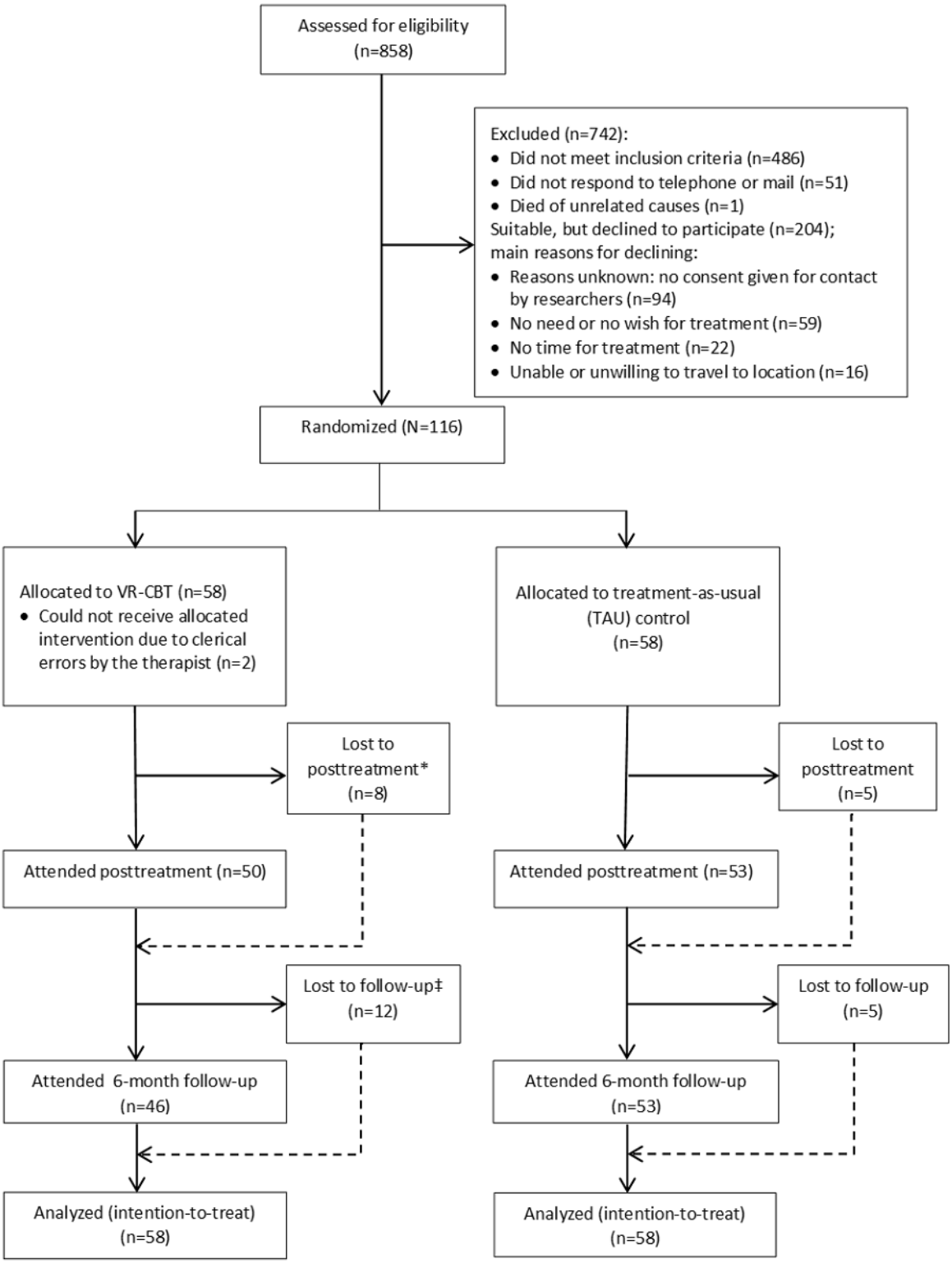


Table 1. Characteristics of the study sample at baseline.

Characteristic	VR-CBT[a] (n=58)	Treatment as usual (TAU) (n=58)
Gender (male), n (%)	40 (69)	42 (72)
Age in years, mean (SD)	36.5 (9.8)	39.5 (10.0)
Non-Dutch origin, n (%)	15 (26)	25 (43)
Education, n (%)		
No education or primary	16 (28)	16 (28)
Vocational	18 (31)	24 (41)
Secondary	9 (16)	9 (16)
Higher	15 (26)	9 (16)
DSM-IV[b]diagnosis		
Schizophrenia, n (%)	46 (79)	49 (85)
Schizoaffective disorder, n (%)	1 (2)	5 (9)
Delusional disorder, n (%)	1 (2)	0 (0)
Psychotic disorder (not otherwise specified), n (%)	10 (17)	4 (7)
Duration of illness in years, mean (SD)	13.3 (10.6)	14.9 (9.5)
Medication use		
Antipsychotics, n (%)	54 (93)	57 (98)
Olanzapine equivalent of prescribed antipsychotic medication (mg/day), mean (SD)	10.5 (6.8)	11.0 (8.3)
Antidepressants, n (%)	15 (26)	17 (29)
Paid work, n (%)	8 (14)	5 (9)
Safety behaviors, mean (SD)	28.8 (14.2)	21.1 (16.0)

[a]VR-CBT: virtual reality-based cognitive behavioral therapy.

[b]DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition.

Table 2. Average per-participant costs per 3-month period in Euros for the year 2015 and average outcomes by measurement and condition.

Costs and outcomes	Baseline		Posttreatment (3 months)		Follow-up (6 months)	
	VR-CBT[a]	TAU[b]	VR-CBT	TAU	VR-CBT	TAU
Costs (€), mean (SD)						
Health care costs	1918 (5178)	1396 (3146)	3031 (3189)	648 (960)	887 (1160)	1039 (2640)
Travel costs	31 (23)	29 (26)	60 (34)	23 (15)	28 (22)	24 (16)
Productivity loss	553 (2730)	224 (1214)	359 (1205)	214 (1127)	28 (161)	102 (588)
Total (societal) costs	2502 (6246)	1649 (3570)	3076 (3469)	885 (1589)	943 (1185)	1165 (2766)
Outcomes, mean (SD)						
GPTS[c] paranoia (score)	85 (34)	77 (31)	70 (31)	75 (31)	67 (33)	75 (33)
Time spent with others (proportion)	0.416 (0.256)	0.364 (0.266)	0.404 (0.226)	0.323 (0.266)	0.419 (0.209)	0.340 (0.273)
Momentary anxiety (score[d])	2.986 (1.120)	3.259 (1.484)	2.586 (1.089)	3.221 (1.495)	2.645 (1.095)	3.218 (1.388)
Momentary paranoia (score[d])	3.064 (1.393)	3.259 (1.418)	2.714 (1.291)	3.221 (1.518)	2.719 (1.293)	3.218 (1.467)

[a]VR-CBT: virtual reality-based cognitive behavioral therapy.

[b]TAU: treatment as usual.

[c]GPTS: Green's Paranoid Thoughts Scale.

[d]Scores are on a 7-point Likert scale ranging from 1 (not at all) to 7 (very).

Incremental Effects

Time Spent With Others

The treatment response rate regarding the time spent with others was 13 patients out of 58 (22%) in the control group and 24 patients out of 58 (41%) in the experimental group. The baseline-adjusted between-group difference between the response rates (ie, the incremental effect) was 0.23, which was statistically significant (SE=0.076, t_{113}=3.07, 95% CI 0.08-0.38, *P*=.003).

Momentary Anxiety

The treatment response rate with regard to momentary anxiety was 17 patients out of 58 (29%) in the control group and 24 patients out of 58 (41%) in the experimental group. The between-group difference between the treatment response rates (ie, incremental effect) was 0.12, but this difference was not statistically significant (SE=0.089, t_{114}=1.36, 95% CI –0.055 to 0.290, *P*=.18).

Momentary Paranoia

The treatment response rate in momentary GPTS paranoia was 11 patients out of 58 (19%) in the control group and 28 patients out of 58 (48%) in the experimental group. The between-group difference in the response rates was 0.29 and was statistically significant (SE=0.0841, t_{114}=3.48, 95% CI 0.126-0.460, *P*=.001).

Quality-Adjusted Life Years

The SMD of GPTS paranoia was 0.523, which was statistically significant (SE=0.120, t_{114}=4.37, 95% CI 0.285-0.760, *P*<.001). Using Sanderson et al's conversion factor [14] of 0.1835 and taking into account a follow-up period of half a year, this became a QALY gain of 0.048 (0.523 × 0.1835 × 0.5) favoring the VR-CBT condition and this was statistically significant (SE=0.011, t_{114}=4.37, 95% CI 0.026-0.069, *P*<.001).

Incremental Costs

Incremental Health Care Costs

As can be seen in Table 2, the average per-patient health care costs in the TAU group was €1396 at baseline, €648 at posttreatment, and €1039 at follow-up. The average per-patient health care costs in the VR-CBT group was €1918 at baseline, €3031 at posttreatment, and €887 at follow-up. This includes €373.95 per patient for VR-related costs included in the posttreatment costs.

The cumulative costs per patient between baseline and follow-up, including the costs of VR-CBT, were €1686 and €3917 in the TAU and VR-CBT conditions, respectively. The between-group difference was €2231 (€3917–€1686) and was statistically significant (SE=663, t_{114}=3.36, *P*=.001) when not adjusted for the initial cost difference between the conditions at baseline. After adjustment for baseline costs, the incremental health care costs became slightly less at €2170 and retained statistical significance (SE=661, t_{113}=3.28, *P*=.001).

Incremental Costs Stemming From Productivity Losses

A total of 11.2% (13/116) of the participants had paid work. The average costs stemming from productivity losses per person for the TAU group was €224 at baseline, €214 at posttreatment, and €104 at follow-up. The average costs stemming from productivity losses per person for the VR-CBT group was €553 at baseline, €359 posttreatment and €28 at follow-up. The cumulative costs per patient between baseline and follow-up were €317 and €387 in the TAU and VR-CBT conditions, respectively. The between-group difference was €70 (€387–€317) and was not statistically significant (SE=274, t_{114}=–0.26, *P*=.80).

Travel Costs

The average costs stemming from travel per person for the TAU group was €29 at baseline, €23 at posttreatment, and €24 at follow-up. The average travel costs per person for the VR-CBT group was €31 at baseline, €60 at posttreatment, and €28 at follow-up. The cumulative travel costs per patient between baseline and follow-up were €47 and €88 in the TAU and VR-CBT conditions, respectively. The between-group difference was €41 (€88–€47) and was statistically significant (SE=6, t_{114}=–6.73, *P*<.001).

Incremental Costs From the Societal Perspective

The cumulative societal costs per patient between baseline and follow-up were €2050 and €4393 in the TAU and VR-CBT conditions, respectively. The between-group difference was €2343 (€4293–€2050) and was statistically significant (SE=747, t_{114}=–3.14, *P*=.002).

Incremental Cost-Effectiveness Ratios From the Societal Perspective

The mean incremental costs for a positive treatment responder was as follows:

1. Time spent with others: €2343/0.23=€10,069.
2. Momentary anxiety: €2343/0.12=€19,525.
3. Momentary paranoia: €2343/0.29=€8079.
4. The mean incremental cost per QALY: €2343/0.048=€48,868.

Figures 2 to 4 depict the distribution of the 5000 bootstrapped ICERs over the cost-effectiveness plane for each of the social participation measures. Figure 2 depicts time spent with others, the plane illustrates 99.70% of the ICERs fall in the northeast quadrant, indicating that more QALYs are gained for higher costs. Figure 3 depicts momentary anxiety, the plane illustrates 90.74% of the ICERs fall in the northeast quadrant, indicating that more QALYs are gained for higher costs. Figure 4 depicts momentary paranoia, the plane illustrates 99.74% of the ICERs fall in the northeast quadrant, indicating that more QALYs are gained for higher costs.

Figure 5 depicts the distribution of the bootstrapped ICERs over the cost-effectiveness plane, with the vast majority of the ICERs in the northeast quadrant, indicating that more QALYs are gained but for higher costs, while 0.02% of the simulated ICERs fall in the southeast quadrant (ie, QALY gains for lower costs) for the VR-CBT group compared with the TAU group.

Figure 2. Cost-effectiveness plane and willingness to pay (WTP) acceptability curve for time spent with others.

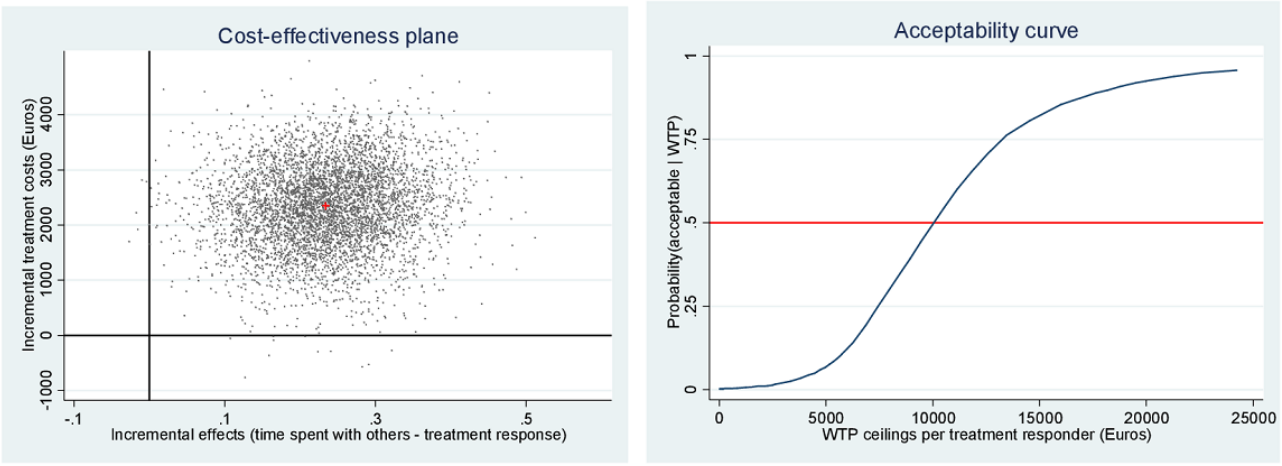


Figure 3. Cost-effectiveness plane and willingness to pay (WTP) acceptability curve for momentary anxiety.

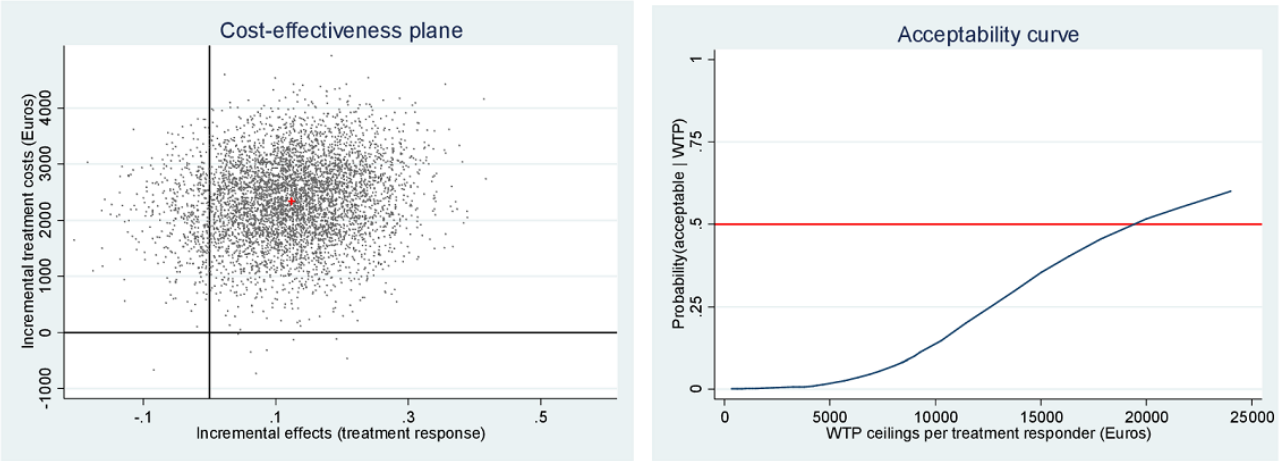


Figure 4. Cost-effectiveness plane and willingness to pay (WTP) acceptability curve for momentary paranoia.

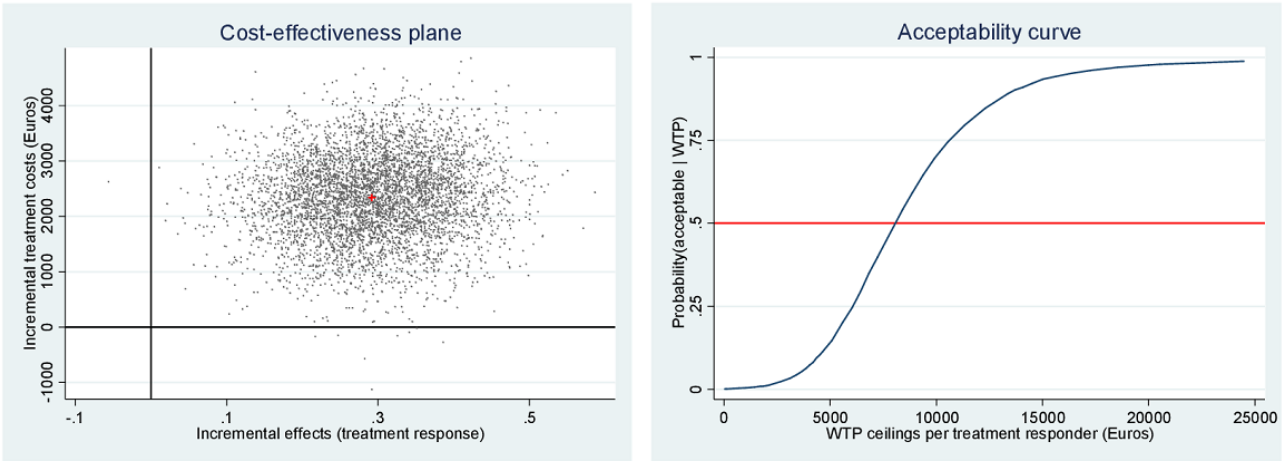
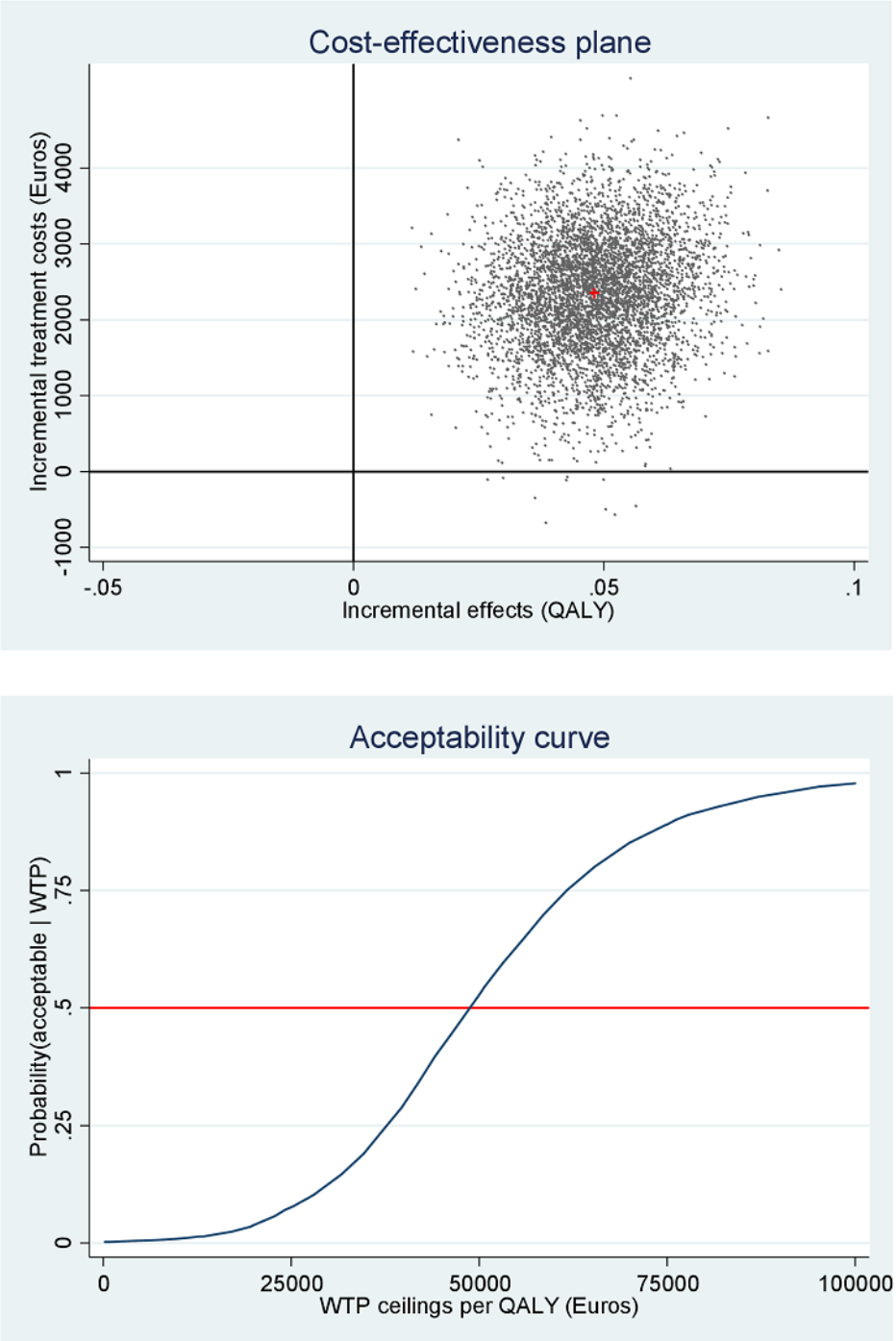


Figure 5. Cost-effectiveness plane and willingness to pay (WTP) acceptability curve for quality-adjusted life year (QALY) gain (costs per QALY gained) after 6 months.



Acceptability

The mean incremental cost per QALY was €48,868. When looking at the acceptability curve in Figure 5, a higher probability that the VR-CBT intervention is deemed cost-effective can also be calculated. For an 80% certainty of cost-effectiveness, the incremental cost for gaining 1 QALY is €66,161, which falls well below the willingness-to-pay ceiling of €80,000 in the Netherlands for a severely disabling condition, such as schizophrenia characterized by paranoid delusions [21].

Looking at the three treatment responses, at 50% probability of being cost-effective, the costs are as mentioned: *time spent with others*, €10,069; *momentary anxiety*, €19,525; and *momentary paranoia*, €8079. Supposing that a decision maker needs an 80% certainty, *time spent with others* will have to be valued at €14,293 per treatment responder; *momentary anxiety* at €50,000; and *momentary paranoia* at €11,342.

Sensitivity Analysis

When including safety behavior at baseline as a covariate, the incremental costs per treatment responder on *time spent with others* became €9136; *momentary anxiety* became €17,535; and *momentary paranoia* became €7219. When including safety behavior at baseline as a covariate, the incremental costs per

QALY gained became €44,597. Overall, incremental costs were somewhat lower when the baseline difference of safety behaviors was included in the analysis.

When including psychiatric admission costs at baseline as a covariate, the incremental costs per treatment responder on *time spent with others* became €9729; *momentary anxiety* became €18,879; and *momentary paranoia* became €7750. When including psychiatric admission at baseline as a covariate, the incremental costs per QALY gained became €47,308. Overall, incremental costs were somewhat lower when the baseline difference of psychiatric admission costs was included in the analysis.

When including both psychiatric admission costs at baseline and safety behavior at baseline as covariates, the incremental costs per treatment responder on *time spent with others* became €8592; *momentary anxiety* became €16,597; and *momentary paranoia* became €6800. When including both psychiatric admission costs and safety behavior at baseline as covariates, the incremental costs per QALY gained became €42,030. Overall, incremental costs were lower when the baseline differences of both safety behaviors and psychiatric admission costs were included in the analysis.

Discussion

Principal Findings

This study aimed to get an impression of short-term cost-effectiveness of VR-CBT for patients with paranoid delusions in comparison to TAU from a societal perspective. Data were collected 6 months after baseline at follow-up. Costs per treatment responder gained were estimated to be between €8079 and €19,525 for different aspects of social participation, with between 90.74% and 99.74% showing improvement. Cost per QALY gained at follow-up was estimated to be €48,868 with 99.98% showing improved QALYs. Sensitivity analyses showed costs to be lower when baseline differences in both safety behavior and psychiatric admission costs were included in the analysis. Costs per treatment responder gained were then estimated to be between €6800 and €16,597, with cost per QALY gained at €42,030.

Results in Context

How much a society values solidarity with people burdened by disease will determine if guidelines are translated to actual treatment of patients. While the VR-CBT treatment condition is more expensive than TAU only, that was to be expected, as the aim was to add to existing treatment. Results show that this addition improves social participation for people with a psychotic disorder suffering from paranoid ideation. We see this improvement for time spent with others, momentary paranoia, momentary anxiety, and paranoid ideation, via the GPTS.

Engaging in psychological therapy is challenging for many patients suffering from paranoid ideation and treatment results vary. There are several aspects that favor VR treatment. Person-specific behavioral exposure is an important part of increasing treatment effect [22], which is exactly what the interactive VR social environments offer. Patients themselves also prefer VR over in vivo exposure treatment [23] and VR improves treatment motivation for patients [24].

Interestingly, during the follow-up we see that the VR-CBT group resulted in decreased health care costs and decreased costs due to productivity loss compared to the TAU-only group. There were no psychiatric admission days at follow-up for the VR-CBT group. To determine whether this was a coincidence or a trend, a much longer follow-up period is needed. Short-term societal costs were between €8079 and €19,525 for a positive treatment response. A disability weight of zero represents no loss of health and a weight of 1 represents health loss equivalent to death [25]. In the Netherlands, the willingness to pay for gaining a QALY ranges between €20,000 and €80,000 but differs per disease [26]. For a severely disabling disease such as schizophrenia, which according to the Global Burden of Disease study 2010 has a disability weight of 0.76, the willingness to pay is €80,000 [21,26]. In this context, the VR-CBT treatment that has an ICER of €48,868 per QALY gained can be regarded as acceptable from the cost-effectiveness point of view.

Limitations

The study has several limitations. First, data were collected only 6 months postbaseline. Any longer-term effects and costs are unknown. There are indications that cost-effectiveness for treatment of psychotic symptoms improves with time [27] as health benefits continue. Second, minimal treatment response was set at a 20% symptom reduction. A 20% symptom reduction after just 8 weeks of therapy is clinically relevant in a patient group with an average duration of illness of 14 years with persistently high problematic isolation. Third, VR-CBT was compared to TAU only. The next step would be to compare VR-CBT directly to CBT, which is the current gold standard, as CBT without VR also results in additional costs to TAU. There are, however, also indications that VR-CBT could have positive results in fewer sessions compared to CBT [9]. Comparing VR-CBT directly to CBT also allows for the study of presumed benefits of VR therapy, such as better engagement and the ecological validity of VR on outcome effects. Such a study comparing VR-CBT and CBT on *time to response* and *costs* is currently ongoing (Netherlands Trial Register number NL7758). A final limitation was that QALYs were not measured directly. As the EQ-5D (European Quality of Life Five Dimension Scale) was not administered, QALYs were calculated using Sanderson et al's conversion factor [14]. Future research needs to include the EQ-5D for direct measurement.

Conclusions

This study found VR-CBT to be cost-effective in the short term from a societal perspective. However, the effect of additional VR-CBT sessions and long-term effects need to be determined while using direct measurement of QALYs.

Acknowledgments

Funding was received from Fonds NutsOhra, Stichting tot Steun VCVGZ.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Direct medical costs. Prices are from 2015 and are in Euros. *8 (sessions) × 2.5 (hours) × 2 (therapists) × 112 (Euros per contact-hour) / 8 (participants) = €560. WRAP: Wellness Recovery Action Plan.
[PNG File , 60 KB - jmir_v22i5e17098_app1.png]

References

1. Jin H, Mosweu I. The societal cost of schizophrenia: A systematic review. Pharmacoeconomics 2017 Jan;35(1):25-42. [doi: 10.1007/s40273-016-0444-6] [Medline: 27557994]
2. Neil AL, Carr VJ, Mihalopoulos C, Mackinnon A, Morgan VA. Costs of psychosis in 2010: Findings from the second Australian National Survey of Psychosis. Aust N Z J Psychiatry 2014 Feb;48(2):169-182. [doi: 10.1177/0004867413500352] [Medline: 24097844]
3. Tajima-Pozo K, de Castro Oller MJ, Lewczuk A, Montañes-Rada F. Understanding the direct and indirect costs of patients with schizophrenia. F1000Res 2015;4:182 [FREE Full text] [doi: 10.12688/f1000research.6699.2] [Medline: 26339474]
4. Moutoussis M, Williams J, Dayan P, Bentall RP. Persecutory delusions and the conditioned avoidance paradigm: Towards an integration of the psychology and biology of paranoia. Cogn Neuropsychiatry 2007 Nov;12(6):495-510. [doi: 10.1080/13546800701566686] [Medline: 17978936]
5. Perkins R, Rinaldi M. Unemployment rates among patients with long-term mental health problems. Psychiatr Bull 2018 Jan 02;26(8):295-298. [doi: 10.1192/pb.26.8.295]
6. Rosenheck R, Leslie D, Keefe R, McEvoy J, Swartz M, Perkins D, CATIE Study Investigators Group. Barriers to employment for people with schizophrenia. Am J Psychiatry 2006 Mar;163(3):411-417. [doi: 10.1176/appi.ajp.163.3.411] [Medline: 16513861]
7. Sibitz I, Amering M, Unger A, Seyringer ME, Bachmann A, Schrank B, et al. The impact of the social network, stigma and empowerment on the quality of life in patients with schizophrenia. Eur Psychiatry 2011 Jan;26(1):28-33. [doi: 10.1016/j.eurpsy.2010.08.010] [Medline: 21036554]
8. Degnan A, Berry K, Sweet D, Abel K, Crossley N, Edge D. Social networks and symptomatic and functional outcomes in schizophrenia: A systematic review and meta-analysis. Soc Psychiatry Psychiatr Epidemiol 2018 Sep;53(9):873-888 [FREE Full text] [doi: 10.1007/s00127-018-1552-8] [Medline: 29951929]
9. Freeman D, Bradley J, Antley A, Bourke E, DeWeever N, Evans N, et al. Virtual reality in the treatment of persecutory delusions: Randomised controlled experimental study testing how to reduce delusional conviction. Br J Psychiatry 2016 Jul;209(1):62-67 [FREE Full text] [doi: 10.1192/bjp.bp.115.176438] [Medline: 27151071]
10. Pot-Kolder RMCA, Geraets CNW, Veling W, van Beilen M, Staring ABP, Gijsman HJ, et al. Virtual-reality-based cognitive behavioural therapy versus waiting list control for paranoid ideation and social avoidance in patients with psychotic disorders: A single-blind randomised controlled trial. Lancet Psychiatry 2018 Mar;5(3):217-226. [doi: 10.1016/S2215-0366(18)30053-1] [Medline: 29429948]
11. Valmaggia LR, Latif L, Kempton MJ, Rus-Calafell M. Virtual reality in the psychological treatment for mental health problems: An systematic review of recent evidence. Psychiatry Res 2016 Feb 28;236:189-195. [doi: 10.1016/j.psychres.2016.01.015] [Medline: 26795129]
12. Pot-Kolder R, Veling W, Geraets C, van der Gaag M. Effect of virtual reality exposure therapy on social participation in people with a psychotic disorder (VRETp): Study protocol for a randomized controlled trial. Trials 2016 Jan 13;17:25 [FREE Full text] [doi: 10.1186/s13063-015-1140-0] [Medline: 26762123]
13. Green CE, Freeman D, Kuipers E, Bebbington P, Fowler D, Dunn G, et al. Measuring ideas of persecution and social reference: The Green et al Paranoid Thought Scales (GPTS). Psychol Med 2008 Jan;38(1):101-111. [doi: 10.1017/S0033291707001638] [Medline: 17903336]
14. Sanderson K, Andrews G, Corry J, Lapsley H. Using the effect size to model change in preference values from descriptive health status. Qual Life Res 2004 Sep;13(7):1255-1264. [doi: 10.1023/B:QURE.0000037482.92757.82] [Medline: 15473504]
15. Bouwmans C, De Jong K, Timman R, Zijlstra-Vlasveld M, Van der Feltz-Cornelis C, Tan Swan S, et al. Feasibility, reliability and validity of a questionnaire on healthcare consumption and productivity loss in patients with a psychiatric disorder (TiC-P). BMC Health Serv Res 2013 Jun 15;13:217 [FREE Full text] [doi: 10.1186/1472-6963-13-217] [Medline: 23768141]

16. van den Brink M, van den Hout WB, Stiggelbout AM, Putter H, van de Velde CJ, Kievit J. Self-reports of health-care utilization: Diary or questionnaire? Int J Technol Assess Health Care 2005;21(3):298-304. [doi: 10.1017/s0266462305050397] [Medline: 16110708]
17. Farmacotherapeutisch Kompas (Pharmaceutical Compass). URL: https://www.farmacotherapeutischkompas.nl/ [accessed 2016-10-01]
18. Swan Tan S, Bouwmans-Frijters CAM, Hakkaart-van Roijen L. Handleiding voor kostenonderzoek: Methoden en referentieprijzen voor economische evaluaties in de gezondheidszorg. TSG 2012 Oct 11;90(6):367-372. [doi: 10.1007/s12508-012-0128-3]
19. Rice DP, Cooper BS. The economic value of human life. Am J Public Health Nations Health 1967 Nov;57(11):1954-1966. [doi: 10.2105/ajph.57.11.1954] [Medline: 6069745]
20. Freeman D, Garety PA, Kuipers E. Persecutory delusions: Developing the understanding of belief maintenance and emotional distress. Psychol Med 2001 Oct;31(7):1293-1306. [doi: 10.1017/s003329170100455x] [Medline: 11681555]
21. Council for Public Health and Health Care (Raad voor de Volksgezondheid en Zorg). Sensible and Sustainable Care (Zinnige en Duurzame Zorg) [document in Dutch]. Zoetermeer, the Netherlands: Council for Public Health and Health Care; 2006 Jun 07. URL: https://www.raadrvs.nl/binaries/raadrvs/documenten/publicaties/2006/06/07/zinnige-en-duurzame-zorg/Advies_-_Zinnige_en_duurzame_zorg.pdf [accessed 2016-10-01]
22. Turner DT, van der Gaag M, Karyotaki E, Cuijpers P. Psychological interventions for psychosis: A meta-analysis of comparative outcome studies. Am J Psychiatry 2014 May;171(5):523-538. [doi: 10.1176/appi.ajp.2013.13081159] [Medline: 24525715]
23. Carl E, Stein AT, Levihn-Coon A, Pogue JR, Rothbaum B, Emmelkamp P, et al. Virtual reality exposure therapy for anxiety and related disorders: A meta-analysis of randomized controlled trials. J Anxiety Disord 2019 Jan;61:27-36. [doi: 10.1016/j.janxdis.2018.08.003] [Medline: 30287083]
24. Rus-Calafell M, Garety P, Sason E, Craig TJ, Valmaggia LR. Virtual reality in the assessment and treatment of psychosis: A systematic review of its utility, acceptability and effectiveness. Psychol Med 2018 Feb;48(3):362-391. [doi: 10.1017/S0033291717001945] [Medline: 28735593]
25. Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: A systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec;380(9859):2197-2223. [doi: 10.1016/S0140-6736(12)61689-4]
26. Zwaap J, Knies S, van der Meijden C, Staal P, van der Heiden L. Cost-Effectiveness in Practice (Kosteneffectiviteit in de Praktijk) [document in Dutch]. Diemen, the Netherlands: Dutch Healthcare Institute (Zorginstituut Nederland); 2015 Jun 26. URL: https://www.zorginstituutnederland.nl/binaries/zinl/documenten/rapport/2015/06/26/kosteneffectiviteit-in-de-praktijk/Kosteneffectiviteit+in+de+praktijk.pdf [accessed 2016-10-01]
27. Ising HK, Lokkerbol J, Rietdijk J, Dragt S, Klaassen RM, Kraan T, et al. Four-year cost-effectiveness of cognitive behavior therapy for preventing first-episode psychosis: The Dutch Early Detection Intervention Evaluation (EDIE-NL) trial. Schizophr Bull 2017 Mar 01;43(2):365-374 [FREE Full text] [doi: 10.1093/schbul/sbw084] [Medline: 27306315]

Abbreviations

3DOF: 3 degrees of freedom
CBT: cognitive behavioral therapy
CEA: cost-effectiveness analysis
CEAC: cost-effectiveness acceptability curve
CHEERS: Consolidated Health Economic Evaluation Reporting Standards
CONSORT: Consolidated Standards of Reporting Trials
CUA: cost-utility analysis
DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition
EQ-5D: European Quality of Life Five Dimension Scale
ESM: ecological sampling method
GPTS: Green's Paranoid Thoughts Scale
ICER: incremental cost-effectiveness ratio
ISRCTN: International Standard Randomised Controlled Trial Number
MICE: multiply imputed chained equations
QALY: quality-adjusted life year
SBQ-PD: Safety Behaviour Questionnaire-Persecutory Delusions
SMD: standard mean difference
SURE: seemingly unrelated regression equations
TAU: treatment as usual
TiC-P: Trimbos Institute and Institute of Medical Technology Assessment Questionnaire for Costs Associated with Psychiatric Illness

VR: virtual reality
VR-CBT: virtual reality-based cognitive behavioral therapy
VU University: Vrije Universiteit

Edited by G Eysenbach; submitted 18.11.19; peer-reviewed by M Rus-Calafell, K Matsumoto; comments to author 15.01.20; revised version received 03.02.20; accepted 06.02.20; published 05.05.20.

Please cite as:
Pot-Kolder R, Veling W, Geraets C, Lokkerbol J, Smit F, Jongeneel A, Ising H, van der Gaag M
Cost-Effectiveness of Virtual Reality Cognitive Behavioral Therapy for Psychosis: Health-Economic Evaluation Within a Randomized Controlled Trial
J Med Internet Res 2020;22(5):e17098
URL: https://www.jmir.org/2020/5/e17098
doi:10.2196/17098
PMID:32369036

Short Paper

Augmented Reality Learning Environment for Basic Life Support and Defibrillation Training: Usability Study

Pier Luigi Ingrassia[1], MD, PhD; Giulia Mormando[2], MD; Eleonora Giudici[3], MD; Francesco Strada[4], MS; Fabio Carfagna[1], MEng; Fabrizio Lamberti[4], PhD; Andrea Bottino[4], PhD

[1]SIMNOVA - Centro di Simulazione in Medicina e Professioni Sanitarie, Università del Piemonte Orientale, Novara, Italy

[2]Department of Medicine, Università di Padova, Padova, Italy

[3]School of Medicine, Università Piemonte Orientale, Novara, Italy

[4]Department of Control and Computer Engineering, Politecnico di Torino, Torino, Italy

Corresponding Author:
Pier Luigi Ingrassia, MD, PhD
SIMNOVA - Centro di Simulazione in Medicina e Professioni Sanitarie
Università del Piemonte Orientale
Via Lanino 1
Novara, 28100
Italy
Phone: 39 3294212083
Email: pierluigi.ingrassia@med.uniupo.it

Abstract

Background: Basic life support (BLS) is crucial in the emergency response system, as sudden cardiac arrest is still a major cause of death worldwide. Unfortunately, only a minority of victims receive cardiopulmonary resuscitation (CPR) from bystanders. In this context, training could be helpful to save more lives, and technology-enhanced BLS simulation is one possible solution.

Objective: The aim of this study is to assess the feasibility and acceptability of our augmented reality (AR) prototype as a tool for BLS training.

Methods: Holo-BLSD is an AR self-instruction training system, in which a standard CPR manikin is "augmented" with an interactive virtual environment that reproduces realistic scenarios. Learners can use natural gestures, body movements, and spoken commands to perform their tasks, with virtual 3D objects anchored to the manikin and the environment. During the experience, users were trained to use the device while being guided through an emergency simulation and, at the end, were asked to complete a survey to assess the feasibility and acceptability of the proposed tool (5-point Likert scale; 1=Strongly Disagree, 5=Strongly Agree).

Results: The system was rated easy to use (mean 4.00, SD 0.94), and the trainees stated that most people would learn to use it very quickly (mean 4.00, SD 0.89). Voice (mean 4.48, SD 0.87), gaze (mean 4.12, SD 0.97), and gesture interaction (mean 3.84, SD 1.14) were judged positively, although some hand gesture recognition errors reduced the feeling of having the right level of control over the system (mean 3.40, SD 1.04).

Conclusions: We found the Holo-BLSD system to be a feasible and acceptable tool for AR BLS training.

(J Med Internet Res 2020;22(5):e14910) doi:10.2196/14910

KEYWORDS

augmented reality; cardiopulmonary resuscitation; automated external defibrillators

Introduction

Sudden cardiac arrest is a major cause of death in adults in developed countries [1]. As such, basic life support (BLS) is a fundamental aspect of the emergency response system. Survival rates are higher when bystanders are able to deliver early cardiopulmonary resuscitation (CPR) and defibrillation [2]. Unfortunately, although laypeople and health care providers are increasingly trained, only a minority of cardiac arrest survivors receive bystander CPR. Therefore, despite major gaps that still exist in the delivery of optimal care [3], training plays a pivotal role in saving lives [4].

The standard approach to BLS training of laypeople involves classroom-based courses consisting of skill demonstrations, hands-on practice, and lectures given by a certified trainer [5]. Simulation with manikins allows trainees to acquire and practice skills without the risk of harming the patient [6]. Debriefing provides an important learning moment in simulation sessions, as it gives participants the opportunity to critically reflect on decisions and actions performed and to learn from mistakes [7]. Although it is still a matter of debate how simulation realism influences learning outcomes [4], the importance of elements of stress and cognitive load in education has been proven, and they should be factored into the instructional design [8]. Self-directed learning systems are an effective alternative to the standard approach, which, according to multiple studies reported by the American Heart Association guidelines, show no statistical difference in learning outcomes compared with instructor-led courses [5].

In the context of self-instruction approaches, emerging technologies allow people to build new cognitive structures [7], and technology-enhanced simulation offers new models for training, which are associated with better knowledge, improved skills acquisition, and a moderate effect on patient outcomes [9]. One such emerging technology is augmented reality (AR), which enhances the user's perception by overlaying virtual objects (or "holograms") on the real-world environment [10]. AR is being applied across various disciplines in health care education including anatomy classes and surgical training [11-13]. In their integrative review, Zhu et al [14] reported that AR could improve health care education by reducing failure rates and improving accuracy.

In this study, we investigated the use of AR as an innovative technology for BLS training. The aim of this study is to assess feasibility and acceptability of Holo-BLSD, our AR prototype tool for CPR training.

Methods

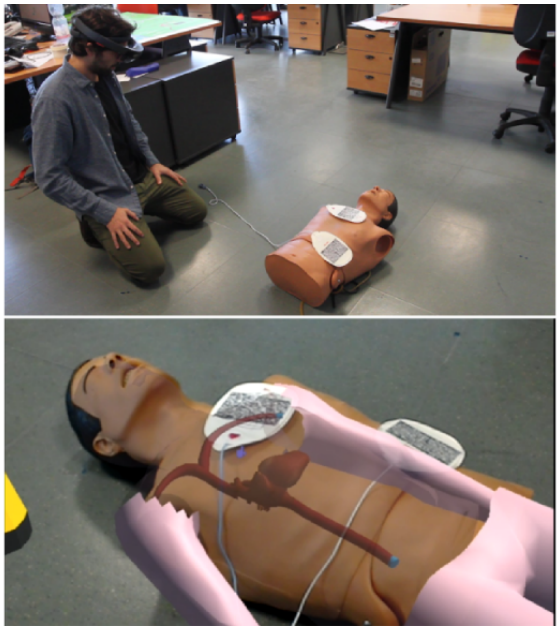
Development

Holo-BLSD was developed jointly by the SIMNOVA simulation center (Novara, Italy) and the Department of Computer Engineering of Politecnico di Torino, in collaboration with Logosnet's e-REAL Immersive Simulation Labs in Lugano, Switzerland. The app uses Microsoft's HoloLens device, a wearable headset for AR experiences, and meets the recent American Heart Association guidelines [2]. Holo-BLSD can be used in different real environments, and its contents can be adapted to where the system is being used: virtual elements can be placed and anchored in the desired position.

Since the majority of potential users are likely to have little or no experience with AR apps or the HoloLens, Holo-BLSD provides a specific training session to help users get acquainted with the system. During interaction training, all interaction types are introduced individually with detailed instructions that include vocal and visual clues.

After the interaction training, users can begin the BLS training. Holo-BLSD guides users step-by-step through the resuscitation procedure of an adult experiencing cardiac arrest. The simulated activities include the following: scene safety, in which any potential hazard should be removed; responsiveness check, performed by shaking and calling the victim; activation of local emergency medical services by using a public phone booth and interacting with a simulated operator; automated external defibrillator (AED) retrieval by asking a witness to find an AED; CPR, using a real manikin torso and superimposing a person's full body using AR (Figure 1); use of AED, once available, by directing pad placement and delivery of shock, if recommended. All generated data are logged, and a feedback sheet can be generated, allowing trainees and instructors to intuitively identify errors and strengths of the performance, thus supporting debriefing sessions and enabling the creation of a library of training events.

Figure 1. A trainee using the HoloLens device (above) and AR displayed in the trainee's field of view (below).



Participants were recruited on a voluntary basis during a national simulation-based training event for medical residents [15]. The selection of subjects represents a convenience sample, and medical background was not a required inclusion criterion. Experiments took place at the SIMNOVA Simulation Center in Novara, Italy.

The experience was divided into three phases, in which volunteers (1) learned how to use the device through the interaction training session, (2) performed a BLS simulation using the Holo-BLSD app and a standard torso manikin, and (3) filled out the evaluation questionnaire.

Software Evaluation

Participants completed a survey of 61 questions divided across six categories (user input, system output, system usability, fidelity of simulation, immersivity, likeability), and graded on a scale from 1 (strongly disagree) to 5 (strongly agree). The complete questionnaire can be found in Multimedia Appendix 1.

The results are presented with relative percentages for each of the five possible grades and summarized as mean and standard deviation. Given the explorative nature of the study, we did not plan any statistical inference tests.

Feasibility

Feasibility was assessed using the user input scale (interaction with gaze and gestures); the system output scale (quality of the display and sound); and the system usability scale [16], which is a standardized tool to evaluate interfaces based on the ISO 9241-400 guidance on ergonomic factors.

Acceptability

The remaining scales of the survey (fidelity, immersion, and likeability) were designed to assess acceptability.

Ethics and Statistical Analysis

Written informed consent was obtained from all participants, and the study results were presented in aggregate with no identifiers. The study was conducted in accordance with the principles of the Declaration of Helsinki. Descriptive statistics were calculated using Microsoft Excel (Version 2003, Microsoft Corporation).

Results

This study involved 26 participants whose characteristics are summarized in Table 1.

Feasibility

Responses to the feasibility scale are shown in Multimedia Appendices 2 and 3 and aggregated data for the system usability scale are presented in Table 2. The cognitive load required to operate the HoloLens was minimal (mean 1.77, SD 0.86) and no high physical effort was required (mean 1.19, SD 0.40). Users felt confident using the software (mean 3.62, SD 1.06). Voice (mean 4.48, SD 0.87), gaze (mean 4.12, SD 0.97), and gesture interaction (mean 3.84, SD 1.14) were evaluated positively. Words and symbols (mean 4.70, SD 0.56) and audio instructions (mean 4.43, SD 0.79) were easy to understand. Users were most critical of the quality of the display (mean 2.45, SD 1.47), although they rated it as appropriate for the function (mean 3.62, SD 1.17). Finally, users rated the system as easy to use (mean 4.00, SD 0.94).

Acceptability

Responses to the acceptability scale are summarized in Multimedia Appendices 4-6. The sensorial information provided by the AR gave participants the impression of physically being in the scenario (mean 3.52, SD 0.95). Users reported that the experience was pleasant (mean 4.13, SD 0.81) and enjoyable (mean 4.65, SD 0.57) and that the virtual contents were realistic (mean 3.74, SD 1.05). Users judged the system as capable of providing real benefit as a training tool (mean 4.22, SD 0.67).

Table 1. Respondents' demographics and characteristics[a].

Characteristic	Participants, n (%)
Gender	
Male	16 (62)
Female	10 (38)
Age group (years)	
20-29	15 (58)
30-39	9 (35)
40-49	0 (0)
50-59	2 (8)
Practice type	
Resident	19 (73)
Physician	3 (12)
Nurse	1 (4)
Other	
Space system engineer	1 (4)
Designer	1 (4)
Secretary	1 (4)
Specialty[b]	
Emergency medicine	7 (32)
Anesthesiology	6 (27)
General surgery	2 (9)
Internal medicine	1 (5)
Pediatrics	1 (5)
Cardiology	1 (5)
Not specified	4 (18)

[a]Percentages may not add up to 100 due to rounding.

[b]Only applicable to residents and physicians.

Table 2. Aggregated data for the system usability scale.

Question	Strongly Disagree (%)	Disagree (%)	Neither (%)	Agree (%)	Strongly Agree (%)	Mean (SD)
System Usability Scale (SUS)						
I think that I would like to use this system frequently	0.0	11.5	23.1	38.5	26.9	3.81 (0.98)
I found the system unnecessarily complex	48.0	36.0	12.0	4.0	0.0	1.72 (0.84)
I thought the system was easy to use	0.0	7.7	19.2	38.5	34.6	4.00 (0.94)
I think that I would need the support of a technical person to be able to use this system	15.4	19.2	30.8	23.1	11.5	2.96 (1.25)
I found the various functions in this system were well integrated	0.0	3.8	30.8	42.3	23.1	3.85 (0.83)
I thought there was too much inconsistency in this system	26.9	34.6	23.1	11.5	3.8	2.31 (1.12)
I would imagine that most people would learn to use this system very quickly	0.0	3.8	26.9	34.6	34.6	4.00 (0.89)
I found the system very cumbersome to use	32.0	24.0	28.0	8.0	8.0	2.36 (1.25)
I felt very confident using the system	0.0	23.1	11.5	46.2	19.2	3.62 (1.06)
I needed to learn a lot of things before I could get going with this system	57.7	26.9	11.5	0.0	3.8	1.65 (0.98)
ISO 9241-400						
The HoloLens device is too bulky or too heavy	19.2	42.3	3.8	26.9	7.7	2.62 (1.30)
The mental effort (concentration) required to operate the device was very high	46.2	34.6	15.4	3.8	0.0	1.77 (0.86)
The physical effort required to operate the device was very high	80.8	19.2	0.0	0.0	0.0	1.19 (0.40)
Arm and hands/fingers fatigue was very high	76.9	19.2	3.8	0.0	0.0	1.27 (0.53)
Eye fatigue was very high	42.3	23.1	19.2	7.7	7.7	2.15 (1.29)
Head fatigue was very high	38.5	38.5	15.4	0.0	7.7	2.00 (1.13)
I would be comfortable using the device for long time	15.4	15.4	30.8	34.6	3.8	2.96 (1.15)

Discussion

Principal Findings

Experts are paying increased attention to the realism and scenario design of CPR training. The physical features of manikins and simulators by themselves are insufficient in suspending learners' disbelief and positively influencing learning outcomes [4]. Despite ongoing advances in resuscitation science, cardiac arrest survival rates remain suboptimal and the educational efficiency of caregivers is still critical, as highlighted in 2003 [17] and 2018 [3]. Education facilitated through technology has been identified as a strategy to improve the effectiveness of BLS training. We developed the Holo-BLSD app using Microsoft's HoloLens technology as an AR self-instruction learning environment for training and assessment of CPR and AED use, by using high definition holograms to immerge trainees in realistic scenarios, and a standard low-cost torso manikin to deliver tactile feedback. In this pilot study, we measured the feasibility and acceptability of our first prototype of the training app.

An excessive cognitive load may impair participants' perceptions and performance, decreasing attention and problem-solving skills [6], and physical effort and fatigue may reduce user enjoyment of the experience. Despite these conditions, users rated the system as easy to use and they judged the learning experience as pleasant and enjoyable. Volunteers reported that the mental effort required to operate the device was minimal. Similarly, users indicated that the head-mounted display was comfortable. Unlike other head-mounted displays, such as the Google Glass used in Chaballout's study [6], no delay between the real and virtual environment was experienced.

Some users noted that ambient light affected the quality of the holograms and made visual instructions difficult to interpret. Users judged the lack of peripheral view, which is known to increase cognitive load [18], as a major limitation. The assembled environment was considered complex enough to be of use, allowing users to focus on the tasks without leaving them disoriented. It is worth noting that trainees found the virtual contents realistic, stating that the system could provide a real benefit as a training tool and help practitioners be more effective.

In this study, we presented preliminary evidence that validated the Holo-BLSD app as a BLS training tool. Additional studies comparing traditional instructor-led training with low and high-fidelity simulation or other models will be useful.

Limitations

The limitations of the study are the small number of participants, which makes it difficult to draw clear conclusions about the benefit of the proposed technology on learning outcomes. Nevertheless, this study was intended to provide a proof of concept and to measure the feasibility and acceptability of AR technology in a life support simulation. To this end, positive user responses indicate that future studies are warranted.

Conclusions

AR is in the early stages of application within health care education, but it has enormous potential. In this study, we presented Holo-BLSD, an AR system for BLS training that offers realistic haptic feedback through a manikin and a virtual scenario that can be easily reconfigured to generate different situations, including extreme and dangerous ones. We found the proposed application to be feasible and acceptable as a tool for self-instruction training. The positive outcomes of this preliminary study make this prototype worthy of future testing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Complete user questionnaire.
[PDF File (Adobe PDF File), 154 KB - jmir_v22i5e14910_app1.pdf]

Multimedia Appendix 2
User input.
[PNG File , 228 KB - jmir_v22i5e14910_app2.png]

Multimedia Appendix 3
System of output.
[PNG File , 135 KB - jmir_v22i5e14910_app3.png]

Multimedia Appendix 4
Fidelity of simulation.
[PNG File , 149 KB - jmir_v22i5e14910_app4.png]

Multimedia Appendix 5
Immersion.
[PNG File , 104 KB - jmir_v22i5e14910_app5.png]

Multimedia Appendix 6
Likeability.
[PNG File , 142 KB - jmir_v22i5e14910_app6.png]

References

1. Yeung J, Okamoto D, Soar J, Perkins GD. AED training and its impact on skill acquisition, retention and performance--a systematic review of alternative training methods. Resuscitation 2011 Jun;82(6):657-664. [doi: 10.1016/j.resuscitation.2011.02.035] [Medline: 21458137]
2. Kleinman ME, Brennan EE, Goldberger ZD, Swor RA, Terry M, Bobrow BJ, et al. Part 5: Adult Basic Life Support and Cardiopulmonary Resuscitation Quality: 2015 American Heart Association Guidelines Update for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. Circulation 2015 Nov 03;132(18 Suppl 2):S414-S435 [FREE Full text] [doi: 10.1161/CIR.0000000000000259] [Medline: 26472993]
3. Cheng A, Nadkarni VM, Mancini MB, Hunt EA, Sinz EH, Merchant RM, American Heart Association Education Science Investigators;on behalf of the American Heart Association Education SciencePrograms Committee, Council on Cardiopulmonary, Critical Care, PerioperativeResuscitation; Council on CardiovascularStroke Nursing;Council on Quality of CareOutcomes Research. Resuscitation Education Science: Educational Strategies to Improve Outcomes From Cardiac Arrest: A Scientific Statement From the American Heart Association. Circulation 2018 Aug 07;138(6):e82-e122. [doi: 10.1161/CIR.0000000000000583] [Medline: 29930020]

4. Hazinski MF, Nolan JP, Aickin R, Bhanji F, Billi JE, Callaway CW, et al. Part 1: Executive Summary: 2015 International Consensus on Cardiopulmonary Resuscitation and Emergency Cardiovascular Care Science With Treatment Recommendations. Circulation 2015 Oct 20;132(16 Suppl 1):S2-39 [FREE Full text] [doi: 10.1161/CIR.0000000000000270] [Medline: 26472854]
5. Bhanji F, Donoghue AJ, Wolff MS, Flores GE, Halamek LP, Berman JM, et al. Part 14: Education: 2015 American Heart Association Guidelines Update for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. Circulation 2015 Nov 03;132(18 Suppl 2):S561-S573. [doi: 10.1161/CIR.0000000000000268] [Medline: 26473002]
6. Chaballout B, Molloy M, Vaughn J, Brisson IR, Shaw R. Feasibility of Augmented Reality in Clinical Simulations: Using Google Glass With Manikins. JMIR Med Educ 2016 Mar 07;2(1):e2 [FREE Full text] [doi: 10.2196/mededu.5159] [Medline: 27731862]
7. Bigi E, Bressan F, Cabrini L, Gasperini C. La simulazione in medicina. Milano: Raffaello Cortina Editore; 2013:978-988.
8. Beckers SK, Biermann H, Sopka S, Skorning M, Brokmann JC, Heussen N, et al. Influence of pre-course assessment using an emotionally activating stimulus with feedback: a pilot study in teaching Basic Life Support. Resuscitation 2012 Mar;83(2):219-226. [doi: 10.1016/j.resuscitation.2011.08.024] [Medline: 21924220]
9. Cook DA, Hatala R, Brydges R, Zendejas B, Szostek JH, Wang AT, et al. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. JAMA 2011 Sep 07;306(9):978-988. [doi: 10.1001/jama.2011.1234] [Medline: 21900138]
10. Hugues O, Fuchs P, Nannipieri O. New Augmented Reality Taxonomy: Technologies and Features of Augmented Environme. In: Furht B, editor. Handbook of augmented reality. New York: Springer; 2011:850.
11. Thomas RG, John NW, Delieu JM. Augmented reality for anatomical education. J Vis Commun Med 2010 Mar;33(1):6-15. [doi: 10.3109/17453050903557359] [Medline: 20297908]
12. Botden SMBI, Jakimowicz JJ. What is going on in augmented reality simulation in laparoscopic surgery? Surg Endosc 2009 Aug;23(8):1693-1700 [FREE Full text] [doi: 10.1007/s00464-008-0144-1] [Medline: 18813987]
13. Ogawa H, Hasegawa S, Tsukada S, Matsubara M. A Pilot Study of Augmented Reality Technology Applied to the Acetabular Cup Placement During Total Hip Arthroplasty. J Arthroplasty 2018 Dec;33(6):1833-1837. [doi: 10.1016/j.arth.2018.01.067] [Medline: 29502961]
14. Zhu E, Hadadgar A, Masiello I, Zary N. Augmented reality in healthcare education: an integrative review. PeerJ 2014;2:e469 [FREE Full text] [doi: 10.7717/peerj.469] [Medline: 25071992]
15. Ingrassia PL, Franc JM, Carenzo L. A novel simulation competition format as an effective instructional tool in post-graduate medical education. Adv Simul (Lond) 2018;3:17 [FREE Full text] [doi: 10.1186/s41077-018-0075-4] [Medline: 30116591]
16. 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]
17. Chamberlain DA, Hazinski MF, European Resuscitation Council, American Heart Association, HeartStroke Foundation of Canada, Australia New Zealand Resuscitation Council, Resuscitation Council of Southern Africa, Consejo Latino-Americano de Resuscitación. Education in resuscitation. Resuscitation 2003 Oct;59(1):11-43. [Medline: 14580733]
18. Baumeister J, Ssin SY, ElSayed NAM, Dorrian J, Webb DP, Walsh JA, et al. Cognitive Cost of Using Augmented Reality Displays. IEEE Trans Vis Comput Graph 2017 Dec;23(11):2378-2388. [doi: 10.1109/TVCG.2017.2735098] [Medline: 28809700]

Abbreviations

AED: automated external defibrillator
AR: augmented reality
BLS: basic life support
CPR: cardiopulmonary resuscitation

Edited by G Eysenbach; submitted 05.06.19; peer-reviewed by J Vaughn, M Padilha; comments to author 11.07.19; revised version received 30.11.19; accepted 15.12.19; published 12.05.20.

Please cite as:
Ingrassia PL, Mormando G, Giudici E, Strada F, Carfagna F, Lamberti F, Bottino A
Augmented Reality Learning Environment for Basic Life Support and Defibrillation Training: Usability Study
J Med Internet Res 2020;22(5):e14910
URL: https://www.jmir.org/2020/5/e14910
doi:10.2196/14910
PMID:32396128

Original Paper

An Informatics Framework to Assess Consumer Health Language Complexity Differences: Proof-of-Concept Study

Biyang Yu[1*], MLIS; Zhe He[1*], MS, PhD; Aiwen Xing[2], BSc; Mia Liza A Lustria[1], MS, PhD

[1]Florida State University, School of Information, Tallahassee, FL, United States

[2]Florida State University, Department of Statistics, Tallahassee, FL, United States

[*]these authors contributed equally

Corresponding Author:
Zhe He, MS, PhD
Florida State University
School of Information
142 Collegiate Loop
Tallahassee, FL, 32306
United States
Phone: 1 850 644 5775
Fax: 1 850 644 9763
Email: zhe@fsu.edu

Abstract

Background: The language gap between health consumers and health professionals has been long recognized as the main hindrance to effective health information comprehension. Although providing health information access in consumer health language (CHL) is widely accepted as the solution to the problem, health consumers are found to have varying health language preferences and proficiencies. To simplify health documents for heterogeneous consumer groups, it is important to quantify how CHLs are different in terms of complexity among various consumer groups.

Objective: This study aimed to propose an informatics framework (consumer health language complexity [CHELC]) to assess the complexity differences of CHL using syntax-level, text-level, term-level, and semantic-level complexity metrics. Specifically, we identified 8 language complexity metrics validated in previous literature and combined them into a 4-faceted framework. Through a rank-based algorithm, we developed unifying scores (CHELC scores [CHELCS]) to quantify syntax-level, text-level, term-level, semantic-level, and overall CHL complexity. We applied CHELCS to compare posts of each individual on online health forums designed for (1) the general public, (2) deaf and hearing-impaired people, and (3) people with autism spectrum disorder (ASD).

Methods: We examined posts with more than 4 sentences of each user from 3 health forums to understand CHL complexity differences among these groups: 12,560 posts from 3756 users in Yahoo! Answers, 25,545 posts from 1623 users in AllDeaf, and 26,484 posts from 2751 users in Wrong Planet. We calculated CHELCS for each user and compared the scores of 3 user groups (ie, deaf and hearing-impaired people, people with ASD, and the public) through 2-sample Kolmogorov-Smirnov tests and analysis of covariance tests.

Results: The results suggest that users in the public forum used more complex CHL, particularly more diverse semantics and more complex health terms compared with users in the ASD and deaf and hearing-impaired user forums. However, between the latter 2 groups, people with ASD used more complex words, and deaf and hearing-impaired users used more complex syntax.

Conclusions: Our results show that the users in 3 online forums had significantly different CHL complexities in different facets. The proposed framework and detailed measurements help to quantify these CHL complexity differences comprehensively. The results emphasize the importance of tailoring health-related content for different consumer groups with varying CHL complexities.

(J Med Internet Res 2020;22(5):e16795) doi:10.2196/16795

KEYWORDS

consumer health informatics; readability; digital divide; health literacy

Introduction

Background

The language gap between laypersons (health consumers) and health care professionals has been long recognized as the main hindrance to effective health communication and health information comprehension [1-3]. When interpreting health documents written mainly in professional language, consumers often depend on their own language to *fill in* the comprehension gap (eg, *depression* vs *depressive disorder*), which might lead to misinterpretation. Accordingly, it has also been widely agreed that health consumers should be given access to resources in their own languages [3-6]. To improve the readability of health-related content for average health consumers, there has been increasing interest in examining consumer health vocabularies [2,7], health readability measurement [8-10], and automated health text simplification approaches [11-14]. Studies on consumer health vocabularies have largely focused on extracting and building a terminology system of lay health terms used by average health consumers [2,7]. Health readability assessments have focused on developing linguistic metrics to quantify the text complexity of health content generated by health experts and professionals [9,13,15,16]. On the basis of the findings in both areas, automated health text simplification usually focuses on simplifying difficult texts with respect to 1 or 2 aspects (eg, medical jargon, long sentences) [1,11,12,14,18,19].

However, without a comprehensive understanding of the complexity difference between professional health language and consumer health language (CHL), current automated simplification approaches are inadequate to accurately determine what needs to be simplified and to what extent they should be simplified. Also, current simplification approaches assume that consumers share the same CHL preferences and that simplifying text to its lowest complexity can satisfy all users. For example, in synonym replacement tasks, researchers typically identify difficult medical words and then replace them with easier synonyms [12,19]. These one-size-fits-all automated simplification approaches ignore the diverse simplification needs of different health customers. Research suggests that consumers with varying health literacy levels have different CHL preferences [20-22]. In addition, contextual and sociocultural factors are found to affect the language preferences of different consumer groups to think, express, and communicate health-related topics [3]. For example, compared with average health consumers, cancer patients would be more familiar with cancer-related professional health terms (eg, genetic predisposition). Another drawback of this one-size-fits-all approach is that simplifying health content by replacing terms with lay alternatives with the lowest complexity may affect information accuracy and may inadvertently increase the length of the text [23]. In other words, an adaptive simplification approach that can balance simplicity, accuracy, and sentence length for user groups with various CHL preferences is ideal.

In this paper, CHL has been defined as a system of vocabularies, expressions, and grammar that is commonly used by a group of health consumers in thinking, expressing, and communicating their health-related topics. CHL complexity is defined as a combined measure of varying linguistic metrics, each of which quantifies the complexity of one linguistic feature of a CHL (eg, semantics, syntax, term). The goal of adaptive health text simplification is to simplify the professional health language used in Web-based health content to match the CHL complexities of targeted consumer groups. To quantify the CHL complexity differences for simplification purposes, the linguistic complexities of CHLs used by various health consumer groups should be investigated. The increasing availability of user-generated Web-based health communications (eg, blogs, online communities, social question and answer [Q&A] websites), provides us with ample opportunities to assess CHL complexity through automated text analysis [2,7,24].

Studies focused on health readability assessment typically quantify the complexity of Web-based health content written by health professionals for health consumers [25-27]. Researchers have developed complexity metrics that utilize a combination of various extracted linguistic features to assess the complexity of Web-based health content [9,13,16]. The metrics utilized in previous literature can be categorized into 4 groups, namely, text-level complexity (eg, syllables per word) [16,28], syntax-level complexity (eg, distributions of parts of speech [POS]) [16,29], term-level complexity (eg, density of professional medical terms) [15,16], and semantic-level complexity (eg, diversity of semantics) [15]. Examining how these linguistic features differ among various CHLs can help us gain a more accurate and comprehensive understanding of CHL complexity.

Objectives

In this proof-of-concept study, we developed an informatics framework (consumer health language complexity [CHELC]) to assess CHL complexity based on existing health text readability metrics and apply this framework to explore complexity differences in CHL in 3 online forums designed for the general public, deaf and hearing-impaired people, and people with autism spectrum disorder (ASD). In previous studies, the latter 2 groups have been found to have relatively low health literacy [30-33], different language use behaviors [34,35], and limited access to adaptive health information services [36]. People with ASD were found to be repetitive and expressive by composing long sentences and words on the Web [35,37,38]. Pollard and Barnett [39] found that even highly educated deaf adults showed significant difficulty in understanding health vocabularies used in the Rapid Estimate of Adult Literacy in Medicine test. In addition, compared with the general population, deaf and hearing-impaired people exhibit significantly lower levels of health literacy and health knowledge [32]. Accordingly, ASD and deaf and hearing-impaired user groups might use less complex CHL, especially less complex health terms in their expressions. Motivated by these observations, in this study, we explore the use of different measures to assess CHL complexity and provide insights for the development of adaptive health text simplification tools to address the needs of various consumer groups.

We formulated 2 research questions (RQs) in this study:

- RQ1: What is the feasibility of using CHELC, which combines text-level, syntax-level, term-level, and semantic-level measures for examining CHL complexity among users in 3 distinct online forums designed for the general public, people with ASD, and deaf and hearing-impaired people?
- RQ2: How do the CHLs of users in online forums designed for the general public, people with ASD, and deaf and hearing-impaired people differ in complexity on the text level, syntax level, term level, and semantic level?

Methods

Consumer Health Language Complexity Measurement Framework

We built CHELC to incorporate a comprehensive array of linguistic complexity metrics developed in previous research. In this framework, we incorporated metrics of text-level, syntax-level, term-level, and semantic-level CHELC scores (CHELCS) to compare various CHLs through a rank-based algorithm. The overall complexity of CHL ($CHELCS_{overall}$) was defined as the average value of 4 complexity scores.

We systematically reviewed the metrics that have been utilized in health readability and complexity assessment studies and comprehensively included credible metrics from all facets of linguistic measures. We performed the search on PubMed using the search terms of *health readability* to retrieve relevant articles and abstracts, which returned 3605 full-text articles to be screened. After excluding duplicates, non-English articles, and articles not about health readability evaluation or assessment, 9 studies with different assessment metrics were identified (Table 1).

Considering the overlap between lay and professional health terms, we proposed to use the ratio of core professional term coverage, which is the percentage of health terms that are in the Systematized Nomenclature of Medicine-Clinical Terms (SNOMED-CT) but not in consumer health vocabulary (CHV). In total, we included 8 metrics for text-level, syntax-level, term-level, and semantic-level complexity measurements in the proposed framework CHELC (Figure 1).

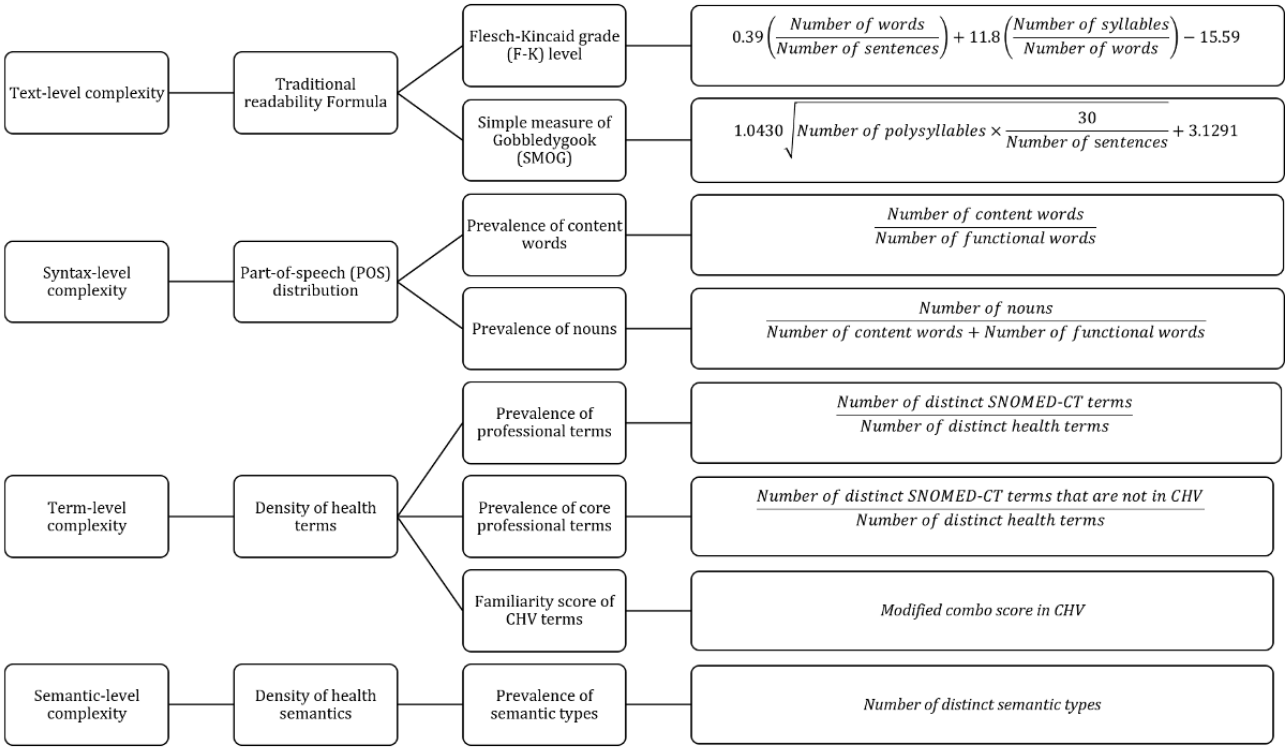
Table 1. Existing metrics for assessing health text complexity.

Health readability measure	Measure specification	Inclusion	Inclusion or exclusion rationale
Text level			
Word length or syllable length [16,28]	Average number of characters (eg, syllables) in a given lexical item	No	Already measured in traditional readability metrics
Sentence length [16,28]	Average number of words in a sentence	No	Already measured in traditional readability metrics
Paragraph length [16,28]	Average number of sentences in a paragraph	No	Not applicable for CHL[a] complexity measure
Traditional readability metrics [10,25,26,40,41]	Flesch-Kincaid grade level, Simple Measure of Gobbledygook, and Gunning fog	Yes	(1) Well-established formulas that are widely utilized in the literature; (2) Combining word, syllable and sentence length; and (3) Flesch-Kincaid grade and Simple Measure of Gobbledygook are the most used readability metrics
Syntax level			
Ratio of content word [15,42]	Ratio of content words (ie, noun, adjective, verb, and adverb) to functional words (ie, pronoun, determiner, preposition, qualifier, conjunction, interjection)	Yes	Indicator for syntax-level complexity measure; validated in previous literature
Ratio of nouns [16,42]	Ratio of nouns to all types of parts of speech	Yes	Indicator for syntax-level complexity measure; validated in previous literature
Term level			
Average familiarity score of CHV[b] [17,28]	Frequency use of each CHV term to the lay people	Yes	Indicator to tell how lay health terms are used in CHL
Coverage in CHV [15]	Ratio of CHV terms of all terms	No	We used the ratio of professional health terms
Coverage in basic medical dictionary [16]	Health terms that are in basic medical dictionaries	No	Not applicable for CHL complexity measure
Coverage in the Unified Medical Language System [15,16]	Ratio of Unified Medical Language System terms	Yes	We utilized the Systematized Nomenclature of Medicine-Clinical Terms as the source of professional health terms
Term overlap ratio [17]	A higher overlap indicates a more cohesive and easier to read text; overlapped terms/all terms in the document	No	Not applicable for CHL complexity measure
Vocabulary size [16]	Distinct word counts in the corpus	No	Not applicable for CHL complexity measure
Style level			
Semantic level			
Diversity of health topics [15]	Ratio of semantic types indicated in the Unified Medical Language System	Yes	Indicator for semantic-level complexity measure; validated in previous literature

[a]CHL: consumer health language.

[b]CHV: consumer health vocabulary.

Figure 1. Consumer health language complexity measurement framework (CHELC).



Text-Level Complexity

Text-level complexity utilizes the length of lexical units (eg, words, sentences, paragraphs) to indicate the lexical complexity of health texts. The unit may change depending on whether the length is applied to words (average number of syllables/characters per word) [16], sentences (average words per sentence) [28], or paragraphs (average sentences per paragraph) [16]. As a commonly used metric, it assumes that longer lexical units require more cognitive loads, thereby making the text more complex. Most studies have utilized one or more readability formulas (eg, the Flesch-Kincaid grade level [F-K] and Simple Measure of Gobbledygook [SMOG]) to assess text-level complexity, in which word length or sentence length are considered in the grade level ranking or level of difficulty of the health texts [10].

For text-level complexity, we applied F-K [43] and SMOG [44] to quantify the text-level complexity of CHL. The F-K formula assigned a grade level to indicate the minimum schooling (grade) readers should have to understand the text. The formula assumes that the higher the average number of syllables and words per sentence there are, the more complex the text is [43]. A grade lower than 5.0 indicates that the text is very easy to comprehend. A grade higher than 12.0 indicates greater difficulty and reading level that requires a college degree or above. Similarly, the SMOG formula considers the number of polysyllabic words [44]. Essentially, the more polysyllabic words, the higher the SMOG score, and the more difficult the texts are.

Syntax-Level Complexity

Syntax-level complexity utilizes POS distribution to evaluate the complexity of health texts [29]. In general, there are 10 commonly used POS types in English, which can be categorized into content words (ie, noun, adjective, verb, adverb) and functional words (ie, pronoun, determiner, preposition, qualifier, conjunction, interjection). Every word in the health text can be assigned a POS tag. A higher proportion of noun words or content words indicates more complex health texts [16]. Accordingly, we calculated the ratio of (1) noun words to all POS words and (2) content words to functional words used by each user. We assume that the higher the ratio is, the more complex the CHL is.

Term-Level Complexity

Term-level complexity focuses on the complexity related to the density of professional or lay terms (eg, *myocardial infarction* vs *heart attack*). According to health readability research, the more professional terms and fewer lay terms there are, the more complex are the health texts [16]. By mapping terms to existing controlled vocabularies, previous studies have typically measured the term-level complexity with the prevalence of professional terms or lay terms [6,15,16]. Other studies have also utilized the familiarity scores of consumer health terms (provided in CHV) and term cohesiveness (ie, distinct word count or overlapped term ratio) to measure the term-level complexity [16,28,].

To assess the term-level complexity of the health text, we first used the text processing and entity recognition tool MetaMap [45] to extract health terms that belong to 84 out of 127 semantic types in the Unified Medical Language System (UMLS, a compendium of over 190 medical controlled vocabularies) that are relevant to biomedicine, health, and nutrition [46,47]. Then we evaluated the density of professional terms and lay terms by mapping our extracted health terms to 2 controlled vocabularies in the UMLS: CHV and SNOMED-CT. CHV contains a collection of lay health concepts and expressions

commonly used by health consumers in their everyday communications [3]. We used the 2015AA version, which includes the latest version of CHV with over 116,324 terms [3]. SNOMED-CT is the world's largest standardized vocabulary of clinical and medical terms mostly used in health information systems such as electronic health records [48-50]. In this study, CHV was used to evaluate the usage of lay health terms, whereas SNOMED-CT terms were referred to as professional terms. We developed the following 3 measures to evaluate term-level complexity:

- Prevalence of professional terms: we used the ratio of professional terms (number of distinct SNOMED-CT terms) to all health-related terms (number of distinct health terms) to measure the density of professional terms used by each user in a health corpus. We assumed that the higher the ratio is, the more complex is the CHL.
- Prevalence of core professional terms: we first excluded CHV terms from SNOMED-CT terms to obtain the core professional terms (professional health terms that are not commonly used by laypersons), and used the ratio of core professional terms to all health-related terms to measure the density of core professional terms used by each user in a health corpus. We assumed that the higher the ratio is, the more complex is the CHL.
- Familiarity score of CHV terms: it refers to the familiarity of each CHV term to laypersons [17]. It is also referred to as the combo score in CHV, which combines frequency score (term difficulty based on its frequency in several large text corpora), context score (term difficulty based on its context), and Concept Unique Identifier score (term difficulty derived from how it is close to well-known easy and difficult concepts in the UMLS). We used a modified combo score that ignores easy words from the Dale-Chall list [17,51]. The higher the score is, the easier the term is. We calculated the average familiarity score of terms written by each user. We assumed that users using more complex CHL have a relatively low average familiarity score for the CHV terms.

Semantic-Level Complexity

Semantic-level complexity refers to the complexity of the diversity of the semantics of health texts. Previous studies have found that if the health text includes more diverse health topics, it is more complex [10]. Operationally, the coverage of semantic types in the UMLS was accounted for semantic-level complexity [47].

We extracted the health terms using MetaMap and counted the average distinct semantic types of the terms used in CHL. We assume that if a user mentioned more distinct semantic types, his or her CHL is more complex.

Consumer Health Language Complexity Scores

We regarded CHL complexity as a 4-faceted variable, which includes metrics related to text-level, syntax-level, term-level, and semantic-level complexity. Each corpus was represented by a vector of 8 metrics for complexity computation. The values of all 8 metrics were generated for every user in the health corpus.

For each metric, the values for users in all health corpora were ranked [52,53] using the same mechanism of Wu et al [16]. In other words, the ranking value for each metric for users was indicated as the complexity differences among users [54,55]. Except for the familiarity score of CHV terms, the higher the metric value is, the more complex the user's health language is. It should be noted that we ranked the familiarity score of CHV terms in reverse order. All the missing values of metrics were replaced by the mean of the corresponding metric.

In this proof-of-concept study, each metric in a facet was regarded to contribute equally to the complexity score of that facet. As there is no agreed-upon definition of health text complexity, each facet has equal weight when calculating the overall complexity score ($CHELCS_{overall}$). The idea of aggregating the metrics is that described by Wu et al [16]. We aggregated the ranks of metrics for each facet using standard aggregate functions with the same weights [56]. Other researchers can use different weights for each metric or facet based on their definitions of CHL complexity.

Let f_{ij} be the j^{th} observed metric value of the i^{th} facet and f'_{ij} be the j^{th} observed metric value of the specific user whose complexity is calculated in the i^{th} facet.

The formula of $CHELCS_{overall}$ for every user in the health corpora was as follows:

We defined r_{ij}, the rank of the j^{th} metric of the i^{th} facet, as the number of users whose f_{ij} is not greater (not smaller for metric *familiarity score of CHV terms*) than $f'_{ij.}$ Note that m represents the number of facets, n_i represents the number of metrics in the i^{th} facet, and N is the total number of users.

We calculated the aggregated rank of the metrics for all facets of CHL complexity. We defined r_{ij}/N as the normalized rank ranging from 0 to 1. Then the aggregated complexity score of the i^{th} facet is calculated as  . The overall complexity score of all facets is calculated as  , which is used to represent the overall CHL complexity of every user. All CHELCS range from 0 to 1, and the higher score means the responding user has more complex CHL complexity in all health corpora.

Data Collection

We utilized CHELC, a complexity measure framework that combines text-level ($CHELCS_{text}$), syntax-level ($CHELCS_{syntax}$), term-level ($CHELCS_{term}$), semantic-level ($CHELCS_{semantic}$), and overall ($CHELCS_{overall}$) complexity scores, to compare the CHLs used in online forums targeting 3 user groups: general public, people with ASD, and deaf and hearing-impaired people. We collected data from various online discussion boards and social media to represent the CHL use of our groups of interest. All 3 data sources in this study were chosen because of their popularity in our interest groups and the convenience of data collection.

We chose AllDeaf [57], a leading online community for deaf and hearing-impaired people who can communicate in English. As of June 2017, AllDeaf had 63,566 members and 114,801 threads. This community has 22 forums in which people can communicate different aspects of everyday life concerns related to deafness, such as sign language, assistive technologies, and health. The majority of the health-related issues are discussed in the forum *Lifestyle, Health, Fitness & Food*. After manually removing the threads that were unrelated to health (eg, food recipes), we retained 1639 threads and 31,006 posts from that forum, which includes health discussions from 2005 to 2016.

Another data source was Wrong Planet [58], which is the main English-language online community developed for people with ASD to discuss everyday life topics. It has 37,350 members and 290,067 threads. Similar to AllDeaf, Wrong Planet has 29 forums. Their users mainly discuss health-related topics in the forum *Health, Fitness & Sports*. After manually removing unrelated threads in that forum, we obtained 2816 threads and 31,194 posts, covering health discussions from 2004 to 2017.

To represent the use of health language by general health consumers, we selected general health discussions in Yahoo! Answers, which is one of the most popular social Q&A sites used by people to discuss health and other life topics. To make the sample size comparable to those collected from AllDeaf and Wrong Planet, we generated a random sample of 8000 questions and their respective answers in the health category, resulting in 34,048 posts from 2009 to 2014.

Data Processing and Analysis

We extracted health-related posts in the 3 forums and calculated CHELCS for each user using text-level ($CHELCS_{text}$), syntax-level ($CHELCS_{syntax}$), term-level ($CHELCS_{term}$), semantic-level ($CHELCS_{semantic}$), and overall ($CHELCS_{overall}$) complexity. As it is not feasible to analyze behavioral patterns for users contributing to few discussions, we only analyzed posts from users who contributed more than 4 sentences per post on average. For the term-level analysis, we only included users who used more than 20 distinct health terms per post. For text- and syntax-level metrics, we generated the scores for each post through a Web-based readability measurement tool [59] and then calculated the complexity score for each user in the 3 corpora using a rank-based algorithm. For the term coverage and semantic analysis, we analyzed the data in MySQL (Oracle Corporation) and Microsoft Excel. We visualized the distributions using $CHELCS_{text}$, $CHELCS_{syntax}$, $CHELCS_{term}$, $CHELCS_{semantic}$, and $CHELCS_{overall}$ for users in each group in Microsoft Excel. Then we employed a 2-sample Kolmogorov-Smirnov test (K-S test) to determine if the CHELCS of the various groups were significantly different. We conducted an analysis of covariance (ANCOVA) to control for possible impacts of sentence number per post on CHELCS when comparing CHL complexity scores of the 3 groups. More detailed comparison results of 3 groups in 8 metrics were presented in Multimedia Appendix 1, and correlations of CHELCS scores were analyzed in Multimedia Appendix 2. K-S test and ANCOVA were performed in R software (The R Foundation for Statistical Computing).

Results

Basic Characteristics of the Corpora

As seen in Table 2, although we extracted similar numbers of posts from the 3 corpora regardless of the number of sentences, the numbers of posts with more than 4 sentences were different among the 3 groups. Compared with the other online forums, Yahoo! Answers had the fewest number of posts, the most threads, and involved the most users, but had the least number of distinct health terms contributed by the average user. This might be because of the differences between specialized online forums that are closed communities and general social Q&A sites that are open to the public [60]. However, the 3 corpora did not have major differences in the number of sentences, sentence lengths, and word lengths, implying that platform differences would not significantly impact the overall CHL used in each community. The 3 user groups shared 68 out of 84 health semantic types in the UMLS.

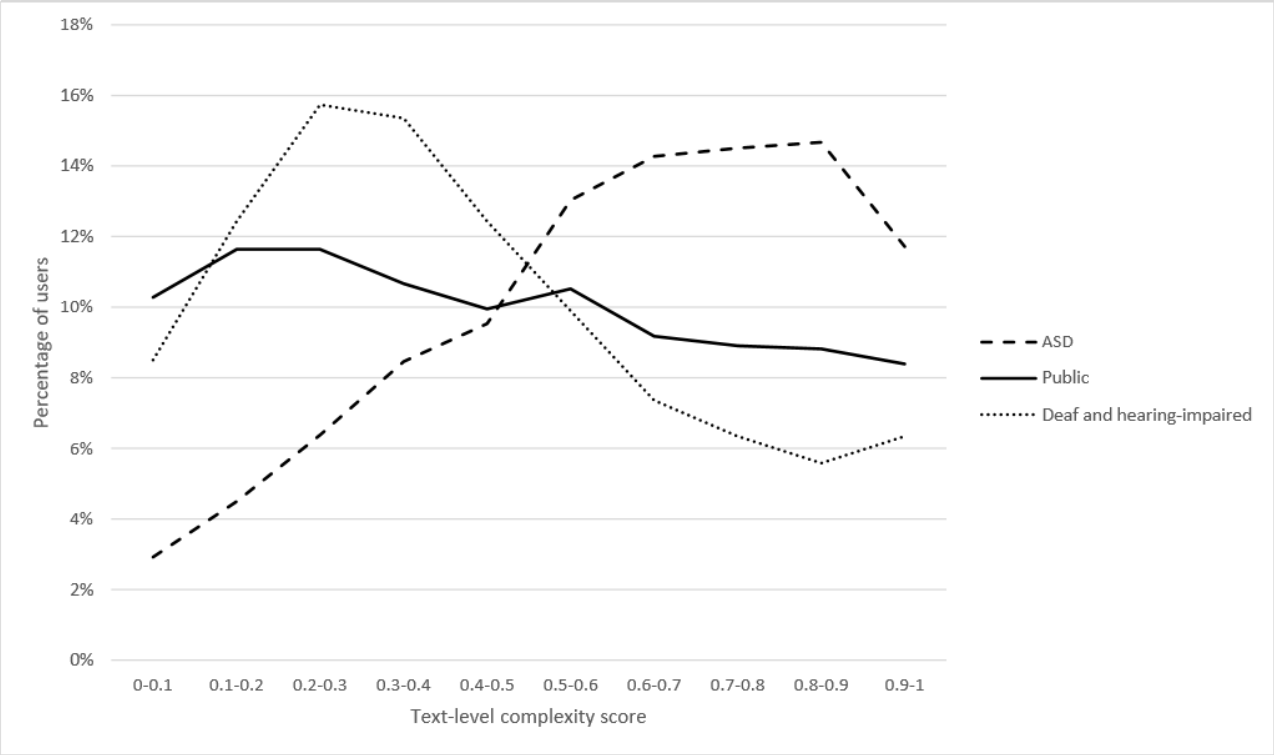
Table 2. Basic textual characteristics of the 3 health corpora.

Basic textual characters	Health corpora		
	AllDeaf (deaf and hearing-impaired people), n	Wrong Planet (people with ASD), n	Yahoo! Answers (general public), n
Number of posts	27,545	26,484	12,560
Number of threads	1623	2751	3756
Number of involved users	788	2978	9544
Average number of sentences per post per user	9.21	9.15	9.63
Average number of words per sentence per user	12.14	13.99	13.09
Average number of syllables per word per user	1.37	1.41	1.35
Average number of letters per word per user	4.14	4.23	4.11
Distinct health terms per user	199.87	91.63	39.09
Mentioned semantics number	71	71	72

Text-Level Complexity

The $CHELCS_{text}$, which ranges from 0 to 1, indicates the text-level complexity ranking of the individual user among all users in the 3 online forums. Figure 2 shows the distribution of text-level complexity scores of users in 3 corpora.

Figure 2. Text-level complexity comparison for users in the 3 health corpora. ASD: autism spectrum disorder.



The 2-sample K-S test results indicate $CHELCS_{text}$ scores of people with ASD, deaf and hearing-impaired people, and the general public were significantly different ($D_{d\text{-}a}$=0.332, $P_{d\text{-}a}$<.001; $D_{d\text{-}p}$=0.108, $P_{d\text{-}p}$<.001; $D_{a\text{-}p}$=0.228, $P_{a\text{-}p}$<.001 [d-a refers to score comparison between *$CHELCS_{text}$ of deaf and hearing-impaired users* and *$CHELCS_{text}$ of users with ASD*; d-p refers to score comparison between *$CHELCS_{text}$ of the deaf and hearing-impaired users* and *$CHELCS_{text}$ of the general public*; a-p refers to score comparison between *$CHELCS_{text}$ of userswith ASD* and *$CHELCS_{text}$ of the general public*]). As seen in Figure 2, most deaf and hearing-impaired users wrote texts with lower complexity, whereas users with ASD used more complex texts in their posts. General public users did not significantly differ in their use of polysyllabic words.

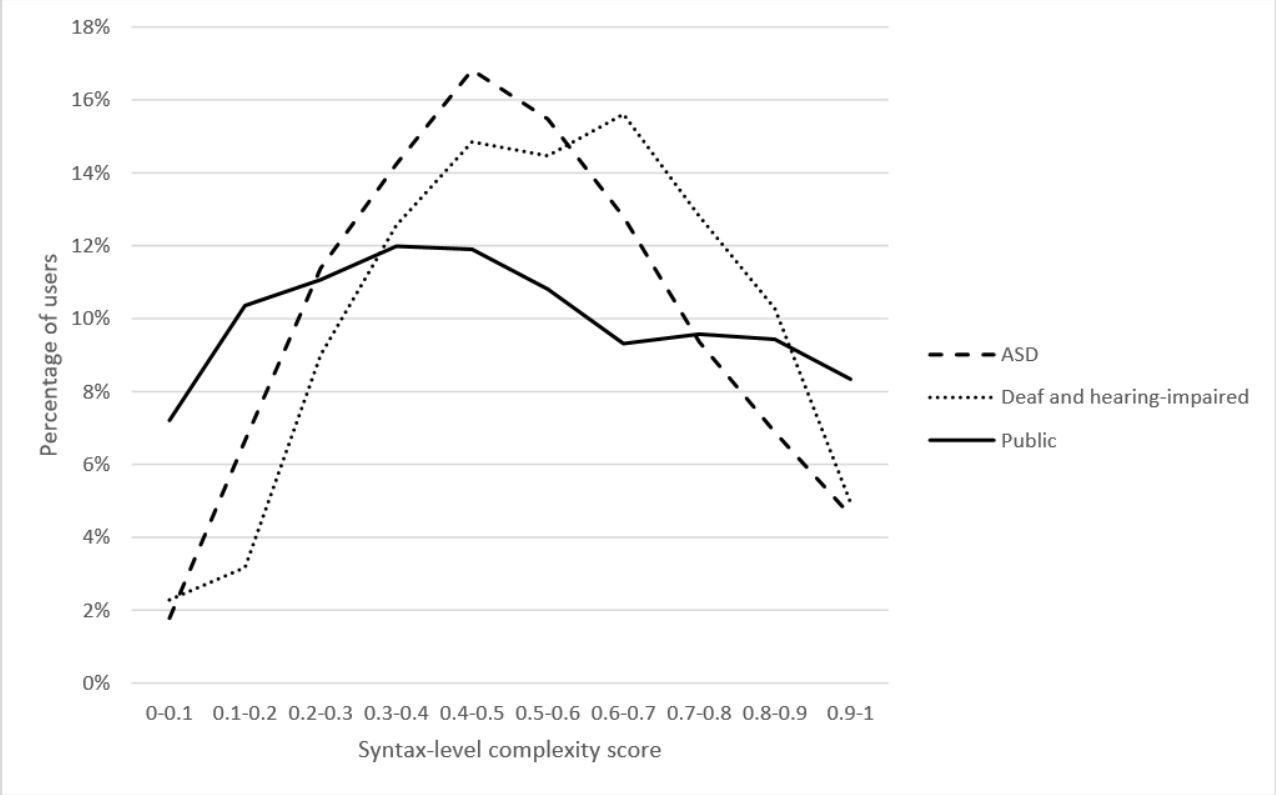
After controlling for the number of sentences per post, the ANCOVA results (F_2=304.5; P<.001) show that users with ASD (mean 0.606) used significantly more complex texts than the other 2 groups (P<.001) and the general public used significantly more complex texts (mean 0.473) than those in the deaf and hearing-impaired group (mean 0.431; P<.001).

Syntax-Level Complexity

The $CHELCS_{syntax}$ indicates complexity ranking related to the prevalence of content words, especially nouns. As seen in Figure 3, the peak $CHELCS_{syntax}$ scores for deaf and hearing-impaired users ranged from 0.6 to 0.7, whereas the peak $CHELCS_{syntax}$ scores for users with ASD ranged from 0.4 to 0.5. Regarding general public users, they did not show a clear syntax complexity preference. The two-sample K-S tests indicate that $CHELCS_{syntax}$ scores were significantly different ($D_{d\text{-}a}$=0.108, $P_{d\text{-}a}$<.001; $D_{d\text{-}p}$=0.153, $P_{d\text{-}p}$<.001; $D_{a\text{-}p}$=0.098, $P_{a\text{-}p}$<.001).

After controlling for the number of sentences per post, the results (F_2=19.206; P<.001) show that deaf and hearing-impaired users used (mean 0.551) significantly more complex syntax than those in the other 2 groups (P<.001), whereas usage of complex syntax was not significantly different between users with ASD (mean 0.506) and the general public (mean 0.494; P=.07).

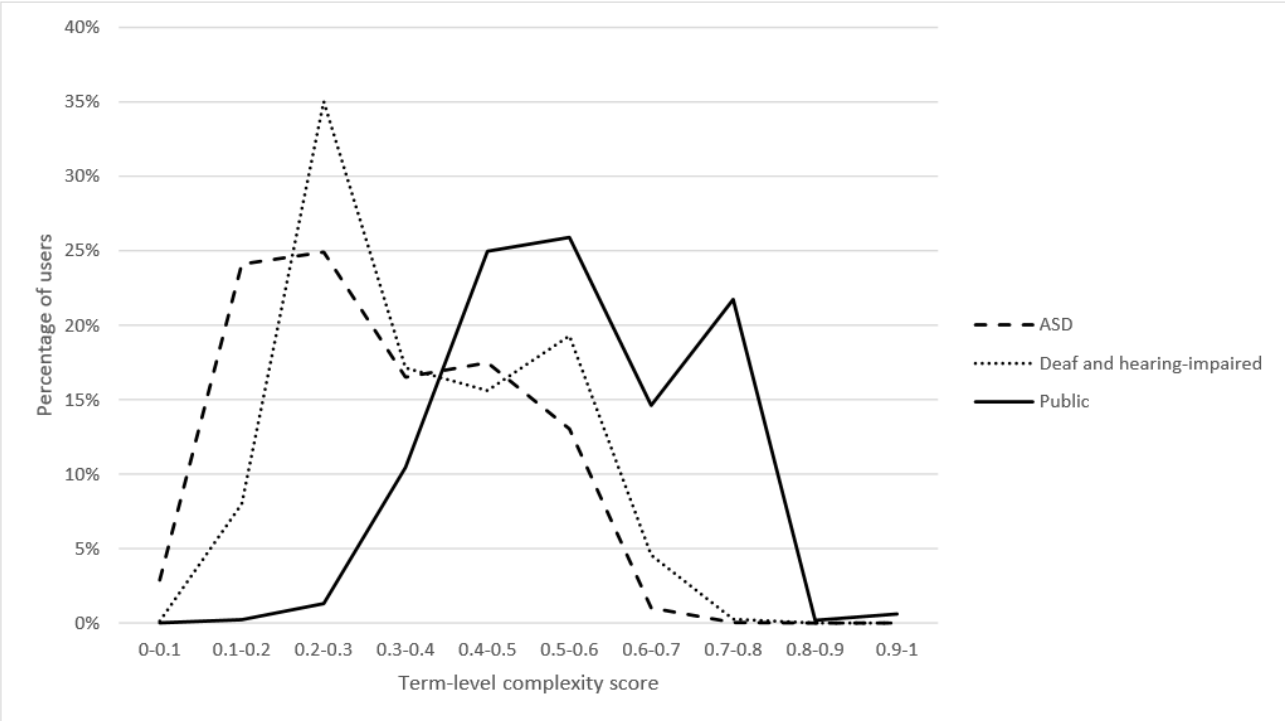
Figure 3. Syntax-level complexity comparison for users in the 3 health corpora. ASD: autism spectrum disorder.



Term-Level Complexity

The $CHELCS_{term}$ focuses on the complexity of the health terms used in each forum. As seen in Figure 4, bimodal distributions were observed in all 3 corpora. Most general public users had relatively higher $CHELCS_{term}$ ranging from 0.2 to 0.9, whereas most users in the other 2 groups had complexity scores lower than 0.7. The two-sample K-S test results indicate that the $CHELCS_{term}$ scores of users with ASD, deaf and hearing-impaired, and general public users were significantly different in the prevalence of professional terms ($D_{d\text{-}a}$=0.208, $P_{d\text{-}a}$=.009; $D_{d\text{-}p}$=0.523, $P_{d\text{-}p}$<.001; $D_{a\text{-}p}$=0.590, $P_{a\text{-}p}$<.001).

Figure 4. Term-level complexity comparison for users in the 3 health corpora. ASD: autism spectrum disorder.

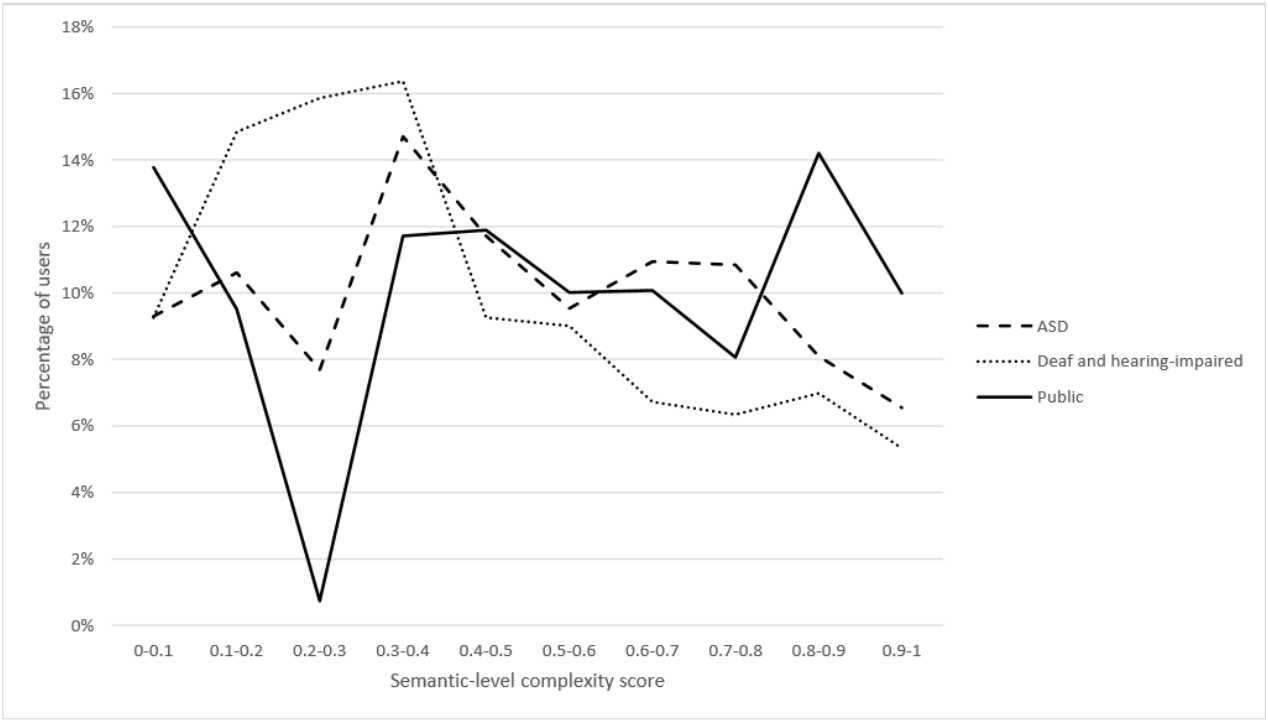


After controlling for the number of sentences per post, the ANCOVA results (F_2=3822.320; P<.001) show that the general public users (mean 0.568) used significantly more complex health terms than those in the other 2 groups (P<.001), and deaf and hearing-impaired users (mean 0.370) used more complex terms than users with ASD (mean 0.316; P<.001).

Semantic-Level Complexity

The $CHELCS_{semantic}$ indicates the diversity of semantic types. Figure 5 shows the distribution of the semantic-level complexity scores in the 3 groups. The two-sample K-S test results indicate that the $CHELCS_{semantic}$ scores for the 3 groups were significantly different ($D_{d\text{-}a}$=0.141, $P_{d\text{-}a}$<.001; $D_{d\text{-}p}$=0.215, $P_{d\text{-}p}$<.001; $D_{a\text{-}p}$=0.116, $P_{a\text{-}p}$<.001). As all health corpora were from social media platforms, the semantics that people utilized might be more influenced by the context than personal health literacy.

Figure 5. Semantic-level complexity comparison for users in the 3 health corpora. ASD: autism spectrum disorder.



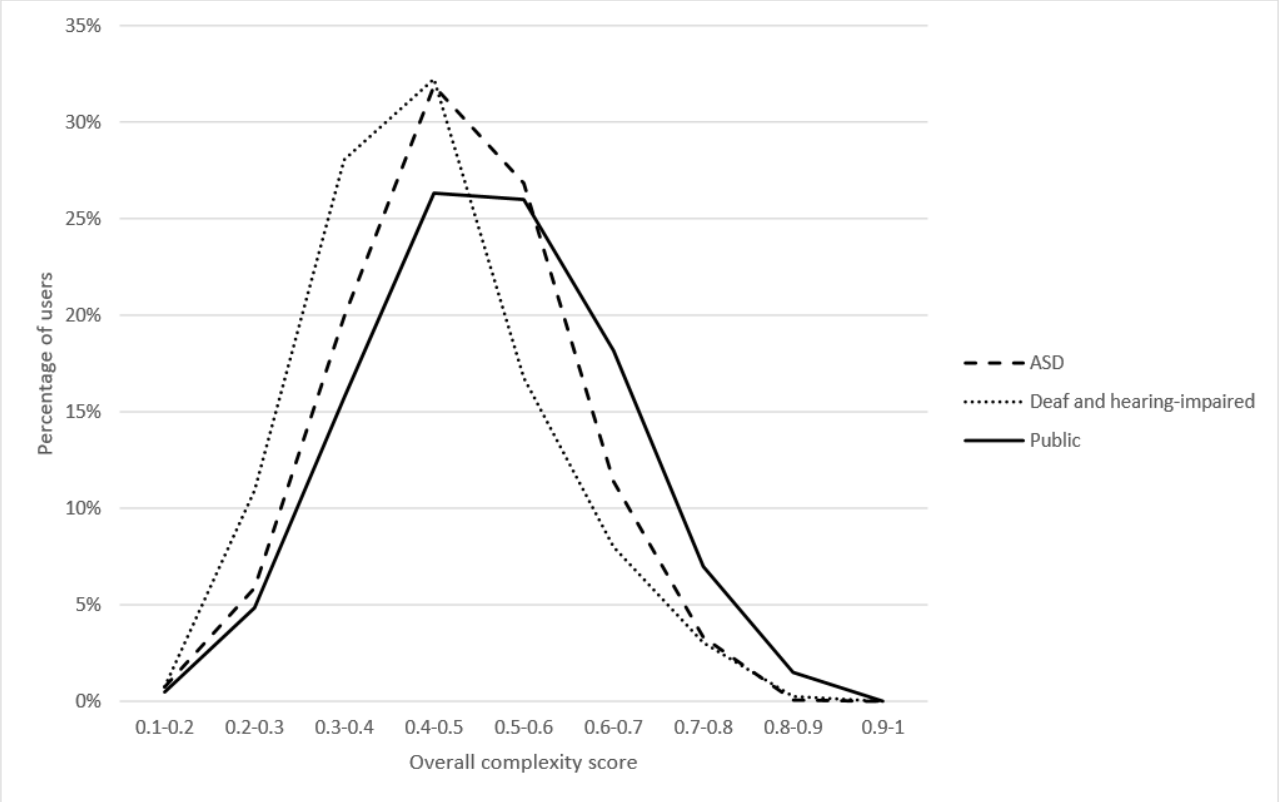
By controlling the number of sentences per post, results (F_2=53.082; P<.001) show that, on average, general public users (mean 0.514) used more semantic types than those in the other 2 groups (P<.001). Users with ASD (mean 0.478) included more semantic types than deaf and hearing-impaired users (mean 0.416; P<.001). In essence, general public users mentioned more diverse health topics than users with ASD and deaf and hearing-impaired users.

Overall Complexity

Figure 6 shows the $CHELCS_{overall}$ for users in the 3 forums. The two-sample K-S test results indicate that the overall CHL complexity scores for users in the 3 corpora were significantly different ($D_{d\text{-}a}$=0.171, $P_{d\text{-}a}$<.001; $D_{d\text{-}p}$=0.250, $P_{d\text{-}p}$<.001; $D_{a\text{-}p}$=0.129, $P_{a\text{-}p}$<.001).

After controlling the number of sentences for each participant, the ANCOVA result (F_2=167.748; P<.001) shows that, on average, general public users (mean 0.512) had more complex CHL than the other 2 groups (P<.001). Users with ASD (mean 0.476) had more complex CHL than deaf and hearing-impaired users (mean 0.442; P<.001).

Figure 6. Overall complexity comparison for users in the 3 health corpora. ASD: autism spectrum disorder.



Discussion

Principal Findings

As health information on the Web often contains medical jargon and complex sentences, general health consumers often find it hard to search for and understand Web-based health information [17]. We argue that health text complexity measurements need to measure the complexity of various CHLs to inform content providers to tailor health information on the Web for health consumers with varying CHL preferences [20,36]. To this end, we developed CHELCS to quantify CHL complexity differences. We applied this measurement to examine CHL complexity differences of health-related posts in 3 online forums targeting the general public, people with ASD, and deaf and hearing-impaired people. In particular, we collected user-generated discussions from 3 online health communities: Yahoo! Answers, Wrong Planet, and AllDeaf. We calculated 8 health readability metrics for each post in the 3 online forums, and calculated text-level ($CHELCS_{text}$), syntax-level ($CHELCS_{syntax}$), term-level ($CHELCS_{term}$), semantic-level ($CHELCS_{semantic}$), and overall ($CHELCS_{overall}$) complexity scores. We then compared the CHL complexity differences for the 3 user groups based on these 5 complexity scores (CHELCS).

The results supported that CHLs of the 3 user groups were significantly different. General public users used more complex health terms and more diverse semantics compared with users with ASD and deaf and hearing-impaired users. Consistent with previous findings, users with ASD used words with more syllables, fewer content or noun words, and less complex health terms [35,38]. Deaf and hearing-impaired users used more content words or nouns, fewer complex words, and less diverse semantics [34,36]. CHELCS results indicated that overall, general public users used more complex CHL than those in the other 2 groups. Overall, the findings from CHELCS measurement were consistent with previous findings of CHL differences among people with ASD, deaf and hearing-impaired people, and public groups.

On the basis of our results, when developing algorithms to simplify health content for different user groups, we need to use more lay health terms for deaf and hearing-impaired users and for users with ASD, less complex words for deaf and hearing-impaired users, and more functional words for users with ASD. For example, as the average F-K grade of MedlinePlus articles is around 8 to 10 [15,16], deaf and hearing-impaired users may need more textual simplifications than the other 2 groups.

To the best of our knowledge, this is the first framework that harnesses consumer-generated textual data to assess the complexity of language that they are comfortable using in their health communications. An understanding of the various CHL complexities of different user groups can provide better insights for the development of adaptive readability assessment tools and adaptive text simplification services.

Limitations

Some limitations should be noted. We could not filter out all the users who are not deaf and hearing impaired or users with ASD, which might affect our findings of the 3 user groups to a certain extent. The data were collected from 3 nontopic–specific health forums. The impact of health topics on text complexity

was not controlled in this exploratory study. For example, CHLs by patients with chronic conditions may be more complex than the average healthy consumers. As the average user contributed little text content in the forums, the findings might not fully depict the language complexity preference of each user. More datasets, such as patient blogs and social media, need to be explored in future studies.

In this proof-of-concept study, the framework CHELC was developed with 8 metrics validated in previous health readability studies to compare CHL complexity differences. Although these metrics have been validated in previous studies, to the best of our knowledge, they have not been used to compare CHLs of different consumer groups. With a lack of research in this field, there is no agreed-upon definition of CHL complexity with respect to different aspects. Therefore, we cannot find a ground truth dataset or standard to validate CHELCS when estimating CHL complexity differences. In this exploratory study, the evaluation of CHELCS was based on previous research findings of the 3 groups in terms of their language complexity preferences. Although our results were consistent with previous findings, this framework and complexity scores are more informative than conclusive. For example, the scores will be different if more metrics are included in this framework, or if the weights of different metrics are defined differently. Also, to more accurately estimate adaptive simplification efforts, it is critical that future studies further assess the CHELCS difference between Web-based consumer health information sources and various CHLs.

Conclusions

The results of this study demonstrate that differences exist among health consumers with respect to the complexity of their language use when discussing health-related topics. A complexity measurement framework (CHELC) and its accompanying scores (CHELCS) were developed to quantify CHL complexity differences among different user groups. Future studies could further apply CHELCS to other datasets from different user groups. Specifically, there is a clear need for the research on understanding CHL complexity differences that translates to adaptive simplification services for different user groups.

Acknowledgments

The authors would like to thank Zhiwei Chen for his help with MetaMap. The authors would also like to thank Dr. Sanghee Oh for sharing with them the data collected from Yahoo! Answers. This project was partially supported by the National Institute on Aging of the National Institutes of Health (NIH) under award number R21AG061431 and the University of Florida Clinical and Translational Science Institute, which is supported in part by the NIH National Center for Advancing Translational Sciences under award number UL1TR001427. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Complexity score of seven metrics in consumer health language complexity scores.
[PDF File (Adobe PDF File), 344 KB - jmir_v22i5e16795_app1.pdf]

Multimedia Appendix 2
Correlations of consumer health language complexity scores in health corpora.
[PDF File (Adobe PDF File), 149 KB - jmir_v22i5e16795_app2.pdf]

References

1. Proulx J, Kandula S, Hill B, Zeng-Treitler Q. Creating Consumer Friendly Health Content: Implementing and Testing a Readability Diagnosis and Enhancement Tool. In: Proceedings of the 2013 46th Hawaii International Conference on System Sciences.: IEEE; 2013 Presented at: HICSS'13; January 7-10, 2013; Wailea, Maui, HI, USA p. 2445-2453. [doi: 10.1109/HICSS.2013.150]
2. Smith CA, Wicks PJ. PatientsLikeMe: Consumer health vocabulary as a folksonomy. AMIA Annu Symp Proc 2008 Nov 6:682-686 [FREE Full text] [Medline: 18999004]
3. Zeng QT, Tse T. Exploring and developing consumer health vocabularies. J Am Med Inform Assoc 2006;13(1):24-29 [FREE Full text] [doi: 10.1197/jamia.M1761] [Medline: 16221948]
4. Lewis D, Brennan PF, McCray AT, Bachman J, Tuttle M, Bachman J. If we build it, they will come: Standardized consumer vocabularies. In: Patel VL, Rogers R, Haux R, editors. MEDINFO 2001. Amsterdam, Netherlands: IOS Press; 2001:1530.
5. Lee KJ. Literature review in computational linguistics issues in the developing field of consumer informatics: finding the right information for consumer's health information need. In: Song M, Wu YF, editors. Handbook of Research on Text and Web Mining Technologies. Hershey, Pennsylvania, USA: IGN Global; 2009:758-765.

6. Keselman A, Logan R, Smith CA, Leroy G, Zeng-Treitler Q. Developing informatics tools and strategies for consumer-centered health communication. J Am Med Inform Assoc 2008;15(4):473-483 [FREE Full text] [doi: 10.1197/jamia.M2744] [Medline: 18436895]
7. He Z, Chen Z, Oh S, Hou J, Bian J. Enriching consumer health vocabulary through mining a social Q&A site: A similarity-based approach. J Biomed Inform 2017 May;69:75-85 [FREE Full text] [doi: 10.1016/j.jbi.2017.03.016] [Medline: 28359728]
8. Clauson KA, Zeng-Treitler Q, Kandula S. Readability of patient and health care professional targeted dietary supplement leaflets used for diabetes and chronic fatigue syndrome. J Altern Complement Med 2010 Jan;16(1):119-124 [FREE Full text] [doi: 10.1089/acm.2008.0611] [Medline: 20064017]
9. Kandula S, Zeng-Treitler Q. Creating a gold standard for the readability measurement of health texts. AMIA Annu Symp Proc 2008 Nov 6:353-357 [FREE Full text] [Medline: 18999150]
10. Leroy G, Miller T, Rosemblat G, Browne A. A balanced approach to health information evaluation: A vocabulary-based naïve Bayes classifier and readability formulas. J Am Soc Inf Sci 2008;59(9):1409-1419. [doi: 10.1002/asi.20837]
11. Kandula S, Curtis D, Zeng-Treitler Q. A semantic and syntactic text simplification tool for health content. AMIA Annu Symp Proc 2010 Nov 13;2010:366-370 [FREE Full text] [Medline: 21347002]
12. Abrahamsson E, Forni T, Skeppstedt M, Kvist M. Medical Text Simplification Using Synonym Replacement: Adapting Assessment of Word Difficulty to a Compounding Language. In: Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations. 2014 Presented at: PITR'14; April 26-30, 2014; Gothenburg, Sweden p. 57-65 URL: https://pdfs.semanticscholar.org/ecc0/d5e746cb30d2c4132df760646b18d823631c.pdf [doi: 10.3115/v1/w14-1207]
13. Kauchak D, Leroy G. Moving beyond readability metrics for health-related text simplification. IT Prof 2016;18(3):45-51 [FREE Full text] [doi: 10.1109/MITP.2016.50] [Medline: 27698611]
14. Leroy G, Endicott JE, Kauchak D, Mouradi O, Just M. User evaluation of the effects of a text simplification algorithm using term familiarity on perception, understanding, learning, and information retention. J Med Internet Res 2013 Jul 31;15(7):e144 [FREE Full text] [doi: 10.2196/jmir.2569] [Medline: 23903235]
15. Leroy G, Helmreich S, Cowie JR, Miller T, Zheng W. Evaluating online health information: beyond readability formulas. AMIA Annu Symp Proc 2008 Nov 6:394-398 [FREE Full text] [Medline: 18998902]
16. Wu DT, Hanauer DA, Mei Q, Clark PM, An LC, Proulx J, et al. Assessing the readability of ClinicalTrials.gov. J Am Med Inform Assoc 2016 Mar;23(2):269-275 [FREE Full text] [doi: 10.1093/jamia/ocv062] [Medline: 26269536]
17. Keselman A, Tse T, Crowell J, Browne A, Ngo L, Zeng Q. Assessing consumer health vocabulary familiarity: an exploratory study. J Med Internet Res 2007 Mar 14;9(1):e5 [FREE Full text] [doi: 10.2196/jmir.9.1.e5] [Medline: 17478414]
18. Leroy G, Endicott JE. Combining NLP With Evidence-Based Methods to Find Text Metrics Related to Perceived and Actual Text Difficulty. In: Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium. USA: ACM; 2012 Presented at: IHI'12; January 28 - 30, 2012; Miami, Florida, USA p. 749-754. [doi: 10.1145/2110363.2110452]
19. Moen H, Peltonen L, Koivumäki M, Suhonen H, Salakoski T, Ginter F, et al. Improving layman readability of clinical narratives with unsupervised synonym replacement. Stud Health Technol Inform 2018;247:725-729. [Medline: 29678056]
20. Jacobs RJ, Caballero J, Ownby RL, Kane MN. Development of a culturally appropriate computer-delivered tailored Internet-based health literacy intervention for Spanish-dominant Hispanics living with HIV. BMC Med Inform Decis Mak 2014 Nov 30;14:103 [FREE Full text] [doi: 10.1186/s12911-014-0103-9] [Medline: 25433489]
21. Claassen AA, van den Ende CH, Meesters JJ, Pellegrom S, Kaarls-Ohms BM, Vooijs J, et al. How to best distribute written patient education materials among patients with rheumatoid arthritis: a randomized comparison of two strategies. BMC Health Serv Res 2018 Mar 27;18(1):211 [FREE Full text] [doi: 10.1186/s12913-018-3039-4] [Medline: 29580277]
22. Chesser A, Burke A, Reyes J, Rohrberg T. Navigating the digital divide: A systematic review of eHealth literacy in underserved populations in the United States. Inform Health Soc Care 2016;41(1):1-19. [doi: 10.3109/17538157.2014.948171] [Medline: 25710808]
23. Shardlow M. A survey of automated text simplification. SpecialIssue 2014;4(1):58-70 [FREE Full text] [doi: 10.14569/specialissue.2014.040109]
24. Yu B, He Z. Exploratory Textual Analysis of Consumer Health Languages for People Who Are D/deaf and Hard of Hearing. In: 2017 IEEE International Conference on Bioinformatics and Biomedicine.: IEEE; 2017 Presented at: BIBM'17; November 13-16, 2017; Kansas City, MO, USA p. 1288-1291. [doi: 10.1109/bibm.2017.8217846]
25. Gemoets D, Rosemblat G, Tse T, Logan R. Assessing readability of consumer health information: an exploratory study. Stud Health Technol Inform 2004;107(Pt 2):869-873. [Medline: 15360936]
26. Walsh TM, Volsko TA. Readability assessment of internet-based consumer health information. Respir Care 2008 Oct;53(10):1310-1315 [FREE Full text] [Medline: 18811992]
27. de Oliveira GS, Jung M, Mccaffery KJ, McCarthy RJ, Wolf MS. Readability evaluation of internet-based patient education materials related to the anesthesiology field. J Clin Anesth 2015 Aug;27(5):401-405. [doi: 10.1016/j.jclinane.2015.02.005] [Medline: 25912728]
28. Kim H, Goryachev S, Rosemblat G, Browne A, Keselman A, Zeng-Treitler Q. Beyond surface characteristics: a new health text-specific readability measurement. AMIA Annu Symp Proc 2007 Oct 11:418-422 [FREE Full text] [Medline: 18693870]

29. Feng L, Jansche M, Huenerfauth M, Elhadad N. A Comparison of Features for Automatic Readability Assessment. In: Proceedings of the 23rd International Conference on Computational Linguistics. USA: Association for Computational Linguistics; 2010 Presented at: COLING'10; August 23-27, 2010; Beijing, China p. 276-284.
30. McKee MM, Paasche-Orlow MK, Winters PC, Fiscella K, Zazove P, Sen A, et al. Assessing health literacy in deaf American sign language users. J Health Commun 2015;20(Suppl 2):92-100 [FREE Full text] [doi: 10.1080/10810730.2015.1066468] [Medline: 26513036]
31. Zazove P, Meador HE, Reed BD, Gorenflo DW. Deaf persons' English reading levels and associations with epidemiological, educational, and cultural factors. J Health Commun 2013;18(7):760-772. [doi: 10.1080/10810730.2012.743633] [Medline: 23590242]
32. Smith SR, Samar VJ. Dimensions of deaf/hard-of-hearing and hearing adolescents' health literacy and health knowledge. J Health Commun 2016;21(sup2):141-154 [FREE Full text] [doi: 10.1080/10810730.2016.1179368] [Medline: 27548284]
33. Koyama T, Tachimori H, Sawamura K, Koyama A, Naganuma Y, Makino H, et al. Mental health literacy of autism spectrum disorders in the Japanese general population. Soc Psychiatry Psychiatr Epidemiol 2009 Aug;44(8):651-657. [doi: 10.1007/s00127-008-0485-z] [Medline: 19096742]
34. Chung JW, Min HJ, Kim J, Park JC. Enhancing Readability of Web Documents by Text Augmentation for Deaf People. In: Proceedings of the 3rd International Conference on Web Intelligence, Mining and Semantics. USA: ACM; 2013 Presented at: WIMS'13; June 12-14, 2013; Madrid, Spain p. 1-10. [doi: 10.1145/2479787.2479808]
35. Gabriels RL, Cuccaro ML, Hill DE, Ivers BJ, Goldson E. Repetitive behaviors in autism: relationships with associated clinical features. Res Dev Disabil 2005;26(2):169-181. [doi: 10.1016/j.ridd.2004.05.003] [Medline: 15590247]
36. Kushalnagar P, Smith S, Hopper M, Ryan C, Rinkevich M, Kushalnagar R. Making cancer health text on the internet easier to read for deaf people who use American sign language. J Cancer Educ 2018 Feb;33(1):134-140 [FREE Full text] [doi: 10.1007/s13187-016-1059-5] [Medline: 27271268]
37. Turner M. Annotation: Repetitive behaviour in autism: a review of psychological research. J Child Psychol Psychiatry 1999 Sep;40(6):839-849. [doi: 10.1111/1469-7610.00502] [Medline: 10509879]
38. Benford P, Standen P. The internet: a comfortable communication medium for people with Asperger syndrome (AS) and high functioning autism (HFA)? J Assist Technol 2009;3(2):44-53. [doi: 10.1108/17549450200900015]
39. Pollard Jr RQ, Barnett S. Health-related vocabulary knowledge among deaf adults. Rehabil Psychol 2009 May;54(2):182-185. [doi: 10.1037/a0015771] [Medline: 19469608]
40. Ley P, Florio T. The use of readability formulas in health care. Psychol Health Med 1996;1(1):7-28. [doi: 10.1080/13548509608400003]
41. Zheng J, Yu H. Readability formulas and user perceptions of electronic health records difficulty: A corpus study. J Med Internet Res 2017 Mar 2;19(3):e59 [FREE Full text] [doi: 10.2196/jmir.6962] [Medline: 28254738]
42. Leroy G, Helmreich S, Cowie JR. The Effects of Linguistic Features and Evaluation Perspective on Perceived Difficulty of Medical Text. In: Proceedings of the 2010 43rd Hawaii International Conference on System Sciences.: IEEE; 2010 Presented at: HICSS'10; January 5-8, 2010; Honolulu, HI, USA p. 1-10 URL: https://ieeexplore.ieee.org/document/5428381 [doi: 10.1109/hicss.2010.374]
43. Flesch RF. How to Write Plain English: A Book for Lawyers and Consumers : With 60 Before-And-After Translations from Legalese. New York, New York, USA: HarperCollins; 1979.
44. McLaughlin GH. SMOG grading: A new readability formula. J Read 1969;12(8):639-646 [FREE Full text]
45. Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. Proc AMIA Symp 2001:17-21 [FREE Full text] [Medline: 11825149]
46. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004 Jan 1;32(Database issue):D267-D270 [FREE Full text] [doi: 10.1093/nar/gkh061] [Medline: 14681409]
47. McCray AT, Nelson SJ. The representation of meaning in the UMLS. Methods Inf Med 1995 Mar;34(1-2):193-201. [doi: 10.1055/s-0038-1634592] [Medline: 9082131]
48. Donnelly K. SNOMED-CT: The advanced terminology and coding system for eHealth. Stud Health Technol Inform 2006;121:279-290. [Medline: 17095826]
49. Agrawal A, He Z, Perl Y, Wei D, Halper M, Elhanan G, et al. The readiness of SNOMED problem list concepts for meaningful use of electronic health records. Artif Intell Med 2013 Jun;58(2):73-80. [doi: 10.1016/j.artmed.2013.03.008] [Medline: 23602702]
50. SNOMED International. 5-Step Briefing URL: http://www.snomed.org/snomed-ct/what-is-snomed-ct [accessed 2018-06-07]
51. National Library of Medicine. UMLS Metathesaurus - CHV (Consumer Health Vocabulary) URL: https://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/CHV/sourcerepresentation.html [accessed 2019-05-03]
52. Ke W, Zhang T, Chen J, Wan F, Ye Q, Han Z. Texture Complexity Based Redundant Regions Ranking for Object Proposal. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops.: IEEE; 2016 Presented at: CVPRW'16; June 26 - July 1, 2016; Las Vegas, NV, USA p. 1083-1091. [doi: 10.1109/cvprw.2016.139]
53. Plumlee MA. The Effect of Information Complexity on Analysts' Use of That Information. Account Rev 2003;78(1):275-296. [doi: 10.2308/accr.2003.78.1.275]

54. Kharlamov E, Giacomelli L, Sherkhonov E, Grau B, Kostylev E, Horrocks I. Ranking, Aggregation, and Reachability in Faceted Search with SemFacet. In: Proceedings of the 16th International Semantic Web Conference. 2017 Presented at: ISWC'17; October 21-25, 2017; Vienna, Austria URL: http://ceur-ws.org/Vol-1963/paper650.pdf
55. Wagner AJ, Ladwig G, Tran T. Browsing-Oriented Semantic Faceted Search. In: Proceedings of the 22nd International Conference on Database and Expert Systems Applications. 2011 Presented at: DEXA'11; August 29 - September 2, 2011; Toulouse, France URL: https://www.aifb.kit.edu/images/8/89/Dexa2011-facetedSearch.pdf
56. Kharlamov E, Giacomelli L, Sherkhonov E, Grau B, Kostylev E, Horrocks I. SemFacet: Making Hard Faceted Search Easier. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. USA: ACM; 2017 Presented at: CIKM'17; November 6 - 10, 2017; Singapore, Singapore p. 2475-2478. [doi: 10.1145/3132847.3133192]
57. Deaf Community. URL: http://www.alldeaf.com/ [accessed 2019-06-01]
58. Wrong Planet. URL: https://wrongplanet.net/ [accessed 2019-06-01]
59. ReadablePro. A Readability Tool, With Extra Power. URL: https://readable.com/ [accessed 2019-03-13]
60. Lerman K, Ghosh R. Arxiv preprints. 2010 Mar 13. Information Contagion: An Empirical Study of the Spread of News on Digg and Twitter Social Networks URL: https://arxiv.org/abs/1003.2664 [accessed 2019-01-05]

Abbreviations

ANCOVA: analysis of covariance
ASD: autism spectrum disorder
CHELC: consumer health language complexity
CHELCS: consumer health language complexity scores
CHL: consumer health language
CHV: consumer health vocabulary
F-K: Flesch-Kincaid grade level
K-S: Kolmogorov-Smirnov
NIH: National Institutes of Health
POS: parts of speech
Q&A: question and answer
SNOMED-CT: Systematized Nomenclature of Medicine-Clinical Terms
UMLS: Unified Medical Language System

Edited by G Eysenbach; submitted 31.10.19; peer-reviewed by D He, X Liu, K Chen; comments to author 16.12.19; revised version received 21.01.20; accepted 21.02.20; published 21.05.20.

Please cite as:
Yu B, He Z, Xing A, Lustria MLA
An Informatics Framework to Assess Consumer Health Language Complexity Differences: Proof-of-Concept Study
J Med Internet Res 2020;22(5):e16795
URL: https://www.jmir.org/2020/5/e16795
doi:10.2196/16795
PMID:32436849

Original Paper

Internet-Based Health Information–Seeking Behavior of Students Aged 12 to 14 Years: Mixed Methods Study

Emanuel Maitz[1]; Katharina Maitz[2], BA, MA; Gerald Sendlhofer[1], PhD; Christina Wolfsberger[3]; Selma Mautner[1], PhD; Lars-Peter Kamolz[1], MSc, Dr med, Prof Dr; Barbara Gasteiger-Klicpera[2], Prof Dr

[1]Division of Plastic, Aesthetic and Reconstructive Surgery, Department of Surgery, Medical University of Graz, Graz, Austria

[2]Inclusive Education Unit, Institute of Professional Development in Education, University of Graz, Graz, Austria

[3]Division of Neonatology, Department of Pediatrics and Adolescent Medicine, Medical University of Graz, Graz, Austria

Corresponding Author:
Emanuel Maitz
Division of Plastic, Aesthetic and Reconstructive Surgery
Department of Surgery
Medical University of Graz
Auenbruggerplatz 29
Graz, A-8036
Austria
Phone: 43 6604756105
Email: emanuel.maitz@stud.medunigraz.at

Abstract

Background: Many children and adolescents are surrounded by smartphones, tablets, and computers and know how to search the internet for almost any information. However, very few of them know how to select proper information from reliable sources. This can become a problem when health issues are concerned, where it is vital to identify incorrect or misleading information. The competence to critically evaluate digital information on health issues is of increasing importance for adolescents.

Objective: The aim of this study was to assess how children and adolescents rate their internet-based health literacy and how their actual literacy differs from their ratings. In addition, there was a question on how their search performance is related to their self-efficacy. To evaluate these questions, a criteria-based analysis of the quality of the websites they visited was performed. Finally, the possibility to increase their internet-based health literacy in a 3-day workshop was explored.

Methods: A workshop with a focus on health literacy was attended by 14 children and adolescents in an Austrian secondary school. After prior assessments (Culture Fair Intelligence Test, revised German version; Reading Speed and Reading Comprehension Test for Grades 6 to 12, German; electronic health literacy scale [eHEALS]; and General Self-Efficacy Scale, Reversed Version, German), the students were asked to perform an internet-based search on a health-related issue. Browser histories and screenshots of all internet searches were gathered, clustered, and analyzed. After the workshop, the health literacy of the students was assessed again by using the eHEALS.

Results: The 14 students opened a total of 85 homepages, but only eight of these homepages were rated as good or fair by two experts (independent rating) based on specific criteria. The analysis showed that the students judged their own internet-based health literacy much higher than the actual value, and students who had rated themselves better did not visit websites of high quality. Internet-based health literacy correlated significantly with the self-efficacy of the students (r_s=0.794, *P*=.002).

Conclusions: Our study showed that it is possible to draw the attention of students to critical aspects of internet search and to slightly improve their search competence in a workshop. Targeted improvement of health literacy is urgently required, and students need special instruction for this purpose. Further investigations in this area with larger sets of data, which could be feasible with the help of a computer program, are urgently needed.

(J Med Internet Res 2020;22(5):e16281) doi:10.2196/16281

KEYWORDS

internet-based health information–seeking behavior; eHealth literacy; children and adolescents; mixed methods study

Introduction

At present, a huge amount of information is accessible on the internet at almost any time and any place. In the United Kingdom, in 2015, 82.8% of adolescents (mean age 13.67 years) used the internet between 1 and 4 hours a day [1]. In the United States, in 2015, 92% of teenagers (aged 13-17 years) used the internet daily, with only 8% of them using the internet once a week or less [2]. In Austria, in 2019, 94% of Austrian youth (aged 11-14 years) had access to the internet at home and spent on average 94 minutes a day using the internet, and 62% of them searched for information using the internet [3]. This is consistent with the finding that 64% of all young people (aged 16-29 years) in the European Union searched for information using the internet in 2015 [4].

The information seeking process usually starts with a need for information [5], which is often satisfied by using search engines (eg, Google) on the internet [6]. Search engines usually require the user to express an information need by formulating a search query. Search queries submitted to a search engine can be quite different, ranging from keywords to whole questions. A recent study showed that children and young adolescents (aged 11-13 years) mainly formulate phrase- and question-like queries rather than keywords when searching for information regarding a particular task [7]. Moreover, young students were more likely to use questions, and the use of keyword queries increased with an increase in the grade level [7]. In this context, a previous study showed that using question-like queries can be advantageous, at least for unexperienced users like young children who have great difficulties in formulating precise and comprehensive keyword queries [8].

Another challenge for children and adolescents who use search engines is the ranking of search results by search engines, with the placement of paid advertisements first, rather than more reliable sources. Children and young adolescents tend to open the top-ranked results, dismissing lower-ranked results [9,10]. In addition, after opening one of the top results, they frequently continue their search only if they are not satisfied with the obtained information. If the first page clicked provides an answer to their question, they complete their internet-based search without evaluating the source [6,11]. However, evaluating the source can help readers assess the reliability and quality of the provided information [12,13]. Knowledge about the person or organization publishing the information can help to recognize the intention of publication (whether it is professional or for sales or marketing). The publication date or the last time a webpage was updated can help to estimate if the information provided is still relevant or is outdated [14,15]. Therefore, readers should evaluate information from authors or webpages they identify as being appropriate [12,14,16].

Children and adolescents in primary and secondary schools usually do not evaluate the sources and information they read on the internet [14,17-19]. They often only evaluate the length of text or the design of a webpage [20]. When middle school students (aged 10-15 years) are asked to evaluate an internet source, it is usually very difficult for them to determine which person or organization authored the information and to assess the actual expertise of the author [11]. Although ninth grade students have a basic idea of the concept of evaluation criteria, they often have difficulties in applying them to information on the internet [19].

Query formulation and source evaluation skills are especially important in health-related searches, because the majority of internet users have searched for health information on the internet at least once [21-23]. Besides health professionals (eg, doctors and nurses) and family or friends, the internet is among the first three choices for most people who have health-related questions [22]. In the European Union, the proportion of people aged 16 to 74 years seeking health-related information on the internet increased from 24% to 51% between 2007 and 2017 [24]. A survey in 2016 showed that 84% of teenagers in the United States (aged 13-18 years) had obtained health information on the internet previously, with 38% searching the internet for health information once a year and 24% searching at least monthly or more frequently [25]. Search engines (eg, Google) are usually used to search for health information on the internet, whereas health portals are rarely used [6,25].

Although the internet can be a very powerful source of health-related information, many experts criticize its trustworthiness. The quality of internet-based health information is consistently rated as problematic [26-29]. Commercial sites, discussion lists, online support groups, and newsgroups frequently provide poor quality or even false health information. In many cases, this information does not meet safety and quality standards. It rather promotes nonevidence-based treatments, provides advice to generally avoid drug intake, or omits possible drug interactions. Furthermore, the content of such webpages is seldom updated [27].

Children and adolescents are rarely aware of these shortcomings of health information on the internet and are especially easily influenced by false or biased information [14,17-19]. Despite the obvious need for educational interventions in schools targeting competence regarding internet-based health information and systematic source evaluation, the respective skills are not part of the national curriculum in most education systems [20,30]. However, in several countries, including Austria, health education is part of the core curriculum [31,32]. In the last decade, the perception of what health education should contain has slowly changed, drawing the attention of policy makers to the concept of health literacy as a complement to health promotion and disease prevention.

Health literacy is defined by the US Department of Health and Human Services as follows: “The degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions.” [33]. Health literacy means the knowledge and competence of finding and understanding health-related information, as well as dealing with that information. For appropriate handling of health-related internet-based information, a high level of health literacy is essential. Health information–seeking skills (finding information) and source evaluation skills (appraising information to make appropriate decisions) can be considered strongly related to health literacy. In addition to information seeking skills and source evaluation

skills, reading literacy (understanding information) and self-efficacy are important influencing factors for health literacy [11]. The concept of self-efficacy was discovered and described by the sociocognitive action theorist Albert Bandura [34-36] and means "own trust in one's own ability." Self-efficacy regarding information seeking and understanding is considered important to start the information seeking process in the first place [11] and supports a person's inclination to make health-related decisions.

Despite an increased awareness of the importance of health literacy in general and the teaching of health literacy–relevant skills, evidence-based intervention programs to improve health literacy in children and adolescents are rare. Often teachers are not familiar with this concept or do not know how to address health literacy in their lessons. One of the main reasons is the lack of valid data regarding the current internet health information–seeking behavior of children and adolescents, especially in Europe (no data available in Austria).

Therefore, the aim of this mixed methods study was to assess the internet-based health information–seeking behavior of secondary school students (aged 12-14 years) in Austria and focus on the quality of information sources used by this age group. The first goal was to find out how competent children and adolescents are when dealing with health-related information on the internet. Additionally, the strategies they used when searching the internet were observed. Finally, we aimed to find out if a 3-day training on interned-based research competence and reading strategies could enhance the students' health literacy. The following research questions were addressed: (1) How did the students rate their internet-based health literacy? (2) Which websites did they visit to access health-related information and what was the quality of the visited websites (before the students received any training)? (3) How does the students' perceived internet-based health literacy relate to the actual quality of their internet search and the selected sources and how accurate is their assumption of their own health literacy? (4) Is there a relationship between the change process of the students' internet-based health literacy (during a workshop that aimed at improving internet health information–seeking behavior and source evaluation) and their general self-efficacy?

Methods

Set-Up and Participants

From March 20 to 22, 2018, we held a workshop that focused on health literacy in Austrian lower secondary students. The workshop was held by one medical student and one education researcher on 3 consecutive days and comprised 12 class hours in total. It took place in the computer laboratory of the school. As our understanding of health literacy includes internet-based information literacy and reading literacy, we addressed these topics explicitly. The participants of the workshop were between 12 and 14 years old, and all of them went to the same third-grade class of a secondary school in Austria. A total of 14 students (eight girls and six boys) participated, and all of them were born in Austria with German as their native language. The parents and the students themselves provided written consent for participation in the workshop and the study.

Instruments

The control variables *general cognitive abilities* and *reading competence* were assessed prior to the workshop using the Culture Fair Intelligence Test, revised German version [37] and the Reading Speed and Reading Comprehension Test for Grades 6 to 12, German [38], respectively. In addition, we asked students about their habits of internet use.

We used the electronic health (eHealth) literacy scale (eHEALS), which has eight statements (Table 1) and an internal consistency of α=.88, to assess the students' internet-related health literacy (pre- and postintervention; on the first day of the workshop and 2 weeks after the workshop) [39]. The statements were scored on a five-point Likert scale with response options ranging from *strongly agree* to *strongly disagree*. For further computations, we reversed the polarity of the scale so that a higher mean score indicated a higher eHealth literacy.

Further, we assessed the general self-efficacy of the participants with the General Self-Efficacy Scale, Reversed Version, German [40], which consists of 10 statements, with four response options for each statement ranging from *strongly disagree* to *strongly agree*. A higher mean score indicated a higher general self-efficacy with sufficient internal consistency between α=.78 and α=.79 [41].

Procedure

At the beginning of the workshop, we asked the students to perform internet-based research on a health-related issue that was described in short narrative text in the fashion of a vignette with a research task. The topics addressed in the vignette were defined by the researchers and the class teacher. We divided the students into two groups (for didactical reasons, it was easier to work with smaller groups), and each group received one of two vignettes (vignette 1 and vignette 2).

Vignette 1 was as follows: "A friend of yours, his name is Klaus, meets with you to discuss something. So far, he has not talked to anyone about it, but now he wants to ask you for some advice. Klaus tells you that since his childhood, he has been bothered about the appearance of his nose, which he finds way too big and too crooked. He is considering having his nose corrected surgically. Now, he would like to know your opinion. You tell him that you would like to give him some advice, but first you want to know more about rhinoplasty. Try to inform yourself using the internet and make a profound decision about what advice you will give Klaus."

Vignette 2 was as follows: "Imagine you have a friend named Claudia. Claudia has a large dark mole on her left cheek since birth. Making matters worse, this mole has dark bristly hair. Claudia tells you that this spot on her cheek has been bothering her since she can remember. In addition, she is afraid that the mole could become dangerous, as some moles do. Now, she asks you if you think she should get rid of the mole. You explain that you would like to give her some advice, but first want to know more about the condition. Try to inform yourself using the internet and make a profound decision about what advice you will give Claudia."

Before and during their search the students did not receive any input on how to search efficiently on the internet or how to evaluate sources. We instructed them to take screenshots of all their Google searches (the students were most familiar with this search engine) and of all webpages they opened. These data were used to determine how they performed their internet-based search and to discuss their search behavior and their results with them later.

Following their search, the students wrote down the advice they would give to their friend (Klaus or Claudia) and why they would give that advice. Thereafter, we discussed the sources they chose and why they chose them. After this first day, we gathered the browser histories of all students and the screenshots they had taken during their searches, as well as the notes they had made on what advice they would have given their friend. On the second and third days of the workshop, the students received input and performed tasks regarding proper search, reading strategies, source evaluation, and reliable internet sources on health issues.

Data Analysis

We clustered and evaluated the browser history and screenshots the students made on the first day of the workshop in order to analyze their internet search behavior. We assessed all visited websites following the classification system of Pérez et al [42] who suggested to distinguish four stages of information quality according to the following (independently rated) criteria: media quality (eg, *A forum versus a website of the Ministry of Health*), author's motivation (eg, *A surgeon with a private practice* versus *a researcher for evidence-based medicine*), and author's position (eg, *A boulevard journalist* versus *a psychologist with specialization in body perception disorders*). Two researchers rated all visited websites independently and subsequently compared and discussed their ratings. There was consensus regarding all websites, except one. For this website, a third researcher was consulted, and eventually, we were able to assign all of the websites to one of the following categories: (1) *good* (3 points; all three dimensions are reliable); (2) *fair* (2 points; two of the three dimensions are reliable); (3) *poor* (1 point; one of the three dimensions is reliable); and (4) *bad* (0 points; all of the three dimensions are unreliable).

Descriptive statistical analyses were performed using IBM SPSS Statistics 25 (IBM Corp, Armonk, New York, USA) [43].

Results

Participant Findings

All participating students had average IQ values (mean 99.77, SD 8.66). In relation to an age-equivalent norm sample, the average performance was within the norm for this cohort in reading comprehension (mean T value 50.07, SD 8.13), and the students were slightly faster than the norm sample in reading speed (mean T value 54.07, SD 5.62). There was a noticeable variation in reading competence among the students; however, reading competence had no relevant associations with seeking behavior-related variables, the eHEALS score, or self-efficacy. Furthermore, on comparing the reading comprehension score with the search behavior of the students, no systematic patterns were identified. A higher reading competence did not lead to a more sophisticated search or better source evaluation.

All students had access to web-enabled devices, such as tablets, smartphones, and computers, at home and reported searching for information on the internet at least several times a month.

Students' Self-Assessment of Their Internet-Based Health Literacy

The self-assessment of internet-based health literacy before the workshop (Table 1) resulted in a rather high mean overall eHEALS score (range 1-5) of 3.5 (SD 0.7). All students indicated that they knew how to find helpful health resources on the internet (Table 1, Q1). In contrast, their answers regarding the availability of health resources on the internet suggested that they did not know which health resources were available or they had little experience with internet-based health resources so far. A detailed inspection of the eHEALS scores (Table 2) showed a high diversity within the scores of the students. The average scores ranged from below the scale mean (student 1) to nearly the highest possible value (student 12). Furthermore, there were hardly any differences between female and male students.

Table 1. Electronic health literacy scale questions and their scores before the workshop.

eHEALS[a] questions	Results before the workshop	
	Participants, n	Score, median (IQR)
Q1: I know how to find helpful health resources on the internet	10	5 (4-5)
Q2: I know how to use the internet to answer my health questions	13	4 (4-5)
Q3: I know what health resources are available on the internet	13	2 (1.5-3)
Q4: I know where to find helpful health resources on the internet	14	3 (1-4)
Q5: I know how to use the health information I find on the internet to help myself	13	4 (4-5)
Q6: I have the skills I need to evaluate the health resources I find on the internet	14	3.5 (2-4)
Q7: I can tell high quality from low quality health resources on the internet	12	3.5 (2-4)
Q8: I feel confident in using information from the internet to make health decisions	13	4 (3.5-4)

[a]eHEALS: electronic health literacy scale.

Table 2. Electronic health literacy scale scores of the students before the workshop.

Student number (sex)	Mean eHEALS[a] score
1 (F)	2.29
2 (F)	2.63
3 (F)	2.71
4 (M)	3.29
5 (F)	3.38
6 (M)	3.5
7 (M)	3.63
8 (F)	3.75
9 (F)	3.88
10 (F)	4.38
11 (F)	4.38
12 (M)	4.5
13 (M)	N/A[b]
14 (M)	N/A

[a]eHEALS: electronic health literacy scale.

[b]N/A: not applicable; data regarding the eHEALS score are not available.

Quality of the Visited Websites

All participants showed similar internet-based information–seeking behavior no matter how competent they rated themselves in the eHEALS questionnaire.

The participating 14 students entered a total of 44 search requests into Google and visited a total of 85 webpages for the task. Among these webpages, 2 (2%) were rated as *good*, 6 (7%) as *fair*, 37 (44%) as *poor*, and 40 (47%) as *bad* by independent raters. Hence, more than 90% of the webpages were classified as *poor* or *bad*. Almost half of the webpages visited were homepages of doctors with obvious commercial backgrounds. These were paid advertisements on the top of the search results list and were labelled accordingly. Only one student visited a webpage of a university providing evidence-based medical information, and another student visited a high-quality medical internet-based encyclopedia. Both webpages were rated as *good* (Table 3).

Table 3. Visited webpages and ratings.

Webpages	Rating (score)[a]	Total visits, n
A medical specialist portal with a medical internet-based encyclopedia for health professionals	Good (3)	1
Evidence-based medical sites without a direct commercial endeavor	Good (3)	1
Video sharing service	Fair (2)	4
An internet-based encyclopedia after the Wiki-Principle	Fair (2)	2
Homepages of doctors with economic backgrounds (paid ads)	Poor (1)	37
Internet-based magazines	Bad (0)	15
Forums	Bad (0)	8
An internet-based tutorial and how-to page	Bad (0)	7
Economically oriented websites (consumer products)	Bad (0)	4
Alternative medicine or pseudomedically oriented pages	Bad (0)	6

[a]Rating was according to the classification scheme of Pérez et al [42].

Perceived Internet-Based Health Literacy Compared With the Actual Quality of Internet Searches and Selected Sources

The mean number of Google searches with unique queries initiated by the participants was 3.14 (SD 1.99). The number of opened websites ranged from 2 to 16 (Table 4). We found no significant correlation (r_s=0.158, P=.59) between the self-assessed internet-based health literacy of the participants and the actual quality of their internet searches. The results of three students were particularly noticeable. Student 2 rated internet-based health literacy relatively low but performed six

Google searches and visited eight websites with an average quality of 1.0. In contrast, student 12 had the highest eHEALS score and visited 16 websites, but the quality of the websites was *poor* or *bad*. Student 9 had the highest average website quality (1.5), and this student's eHEALS score was relatively high.

Table 4. Electronic health literacy scale scores, searches, and search qualities of the students.

Student number (sex)	eHEALS[a] score before the workshop	Number of Google searches	Number of visited websites	Average quality of the websites	Rating (score)[b], number of websites			
					Good (3)	Fair (2)	Poor (1)	Bad (0)
1 (F)	2.29	3	8	0.13	—	—	1	7
2 (F)	2.63	6	8	1	—	2	4	2
3 (F)	2.71	3	5	0.4	—	—	2	3
4 (M)	3.29	1	3	1	—	—	3	1
5 (F)	3.38	3	7	0.29	—	—	2	5
6 (M)	3.5	3	5	1	—	1	3	1
7 (M)	3.63	2	5	0.4	—	—	2	3
8 (F)	3.75	1	4	0.5	—	—	2	2
9 (F)	3.88	1	4	1.5	1	—	3	—
10 (F)	4.38	1	2	0	—	—	—	2
11 (F)	4.38	3	5	1,2	—	1	4	—
12 (M)	4.5	7	16	0.44	—	—	7	9
13 (M)	N/A[c]	4	7	0.71	1	1	1	4
14 (M)	N/A	6	6	0.67	—	1	3	2

[a]eHEALS: electronic health literacy scale.

[b]Rating was according to the classification scheme of Pérez et al [42].

[c]N/A: not applicable; data regarding the eHEALS score are not available.

Relationship Between the Change Process of the Students' Internet-Based Health Literacy and Their General Self-Efficacy

Regarding the difference in eHEALS scores before and after the workshop, we found that the mean eHEALS score increased slightly from 3.5 (SD 0.7) before the workshop to 4.0 (SD 0.5) after the workshop. However, not all students had a higher eHEALS score after the workshop. Those students who estimated their internet-based health literacy as particularly high before the workshop had lower ratings after the workshop and vice versa. We also found a significant correlation (Spearman rank correlation coefficient) between the eHEALS score before the workshop and the students' general self-efficacy (r_s=0.794, P=.002). Participants with a high self-efficacy rated themselves higher in the eHEALS as compared to those with lower self-efficacy. In contrast, participants with high self-efficacy rated their internet-based health literacy lower and those with low self-efficacy rated their internet-based health literacy higher after the workshop than before the workshop (Table 5).

After the workshop, on average, the participants rated themselves better for eHEALS questions Q2, Q3, Q4, Q6, and Q7, but worse for questions Q1, Q5, and Q8. In particular, the effects of the workshop were noticeable for Q1 ("I know how to find helpful health resources on the internet") and Q3 ("I know what health resources are available on the internet"). The score for Q1 was high before the workshop and slightly lower after the workshop. This question was apparently particularly difficult for the students to answer before the workshop, as only 10 of the 14 students provided an answer. The score for Q3 changed in the opposite direction with an increase in the median (Table 6).

Table 5. Difference in electronic health literacy scale scores before and after the workshop and self-efficacy.

Student number	eHEALS[a] score before the workshop	eHEALS score after the workshop	Difference in the eHEALS score	Self-efficacy score
2	2.63	4.17	1.54	2.33
4	3.29	4.75	1.46	2.3
5	3.38	4.14	0.77	1.78
9	3.88	4	0.13	2.6
7	3.63	3.75	0.13	2.78
6	3.5	3.5	0	2.3
10	4.38	3.75	−0.13	3.8
12	4.5	4.13	−0.38	3.14
11	4.38	4.25	−0.63	3.78
8	3.75	3	−0.75	3.5

[a]eHEALS: electronic health literacy scale.

Table 6. Electronic health literacy scale questions and their median scores before and after the workshop.

eHEALS[a] questions	Results before the workshop		Results after the workshop	
	Participants, n	Score, median (IQR)	Participants, n	Score, median (IQR)
Q1: I know how to find helpful health resources on the internet	10	5 (4-5)	12	4 (4-5)
Q2: I know how to use the internet to answer my health questions	13	4 (4-5)	12	5 (4.25-5)
Q3: I know what health resources are available on the internet	13	2 (1.5-3)	12	4 (2.25-4)
Q4: I know where to find helpful health resources on the internet	14	3 (1-4)	12	4 (3.25-4.75)
Q5: I know how to use the health information I find on the internet to help myself	13	4 (4-5)	12	4 (3-4.75)
Q6: I have the skills I need to evaluate the health resources I find on the internet	14	3.5 (2-4)	12	4 (4-5)
Q7: I can tell high quality from low quality health resources on the internet	12	3.5 (2-4)	10	4 (3-5)
Q8: I feel confident in using information from the internet to make health decisions	13	4 (3.5-4)	11	3 (2-4)

[a]eHEALS: electronic health literacy scale.

Students' Statements Regarding Their Learning Experience During the Workshop

At the final interview after the workshop, the students acknowledged the usefulness of the workshop and the importance of critically reviewing information on the internet. For example, one student described the main insights as follows:

> *The search results […] there are mostly ads on the first page and that they only present positive things on their webpages. And only positive things, therefore no negative things. Therefore, you should use general pages that have no special intention, where you can look for advantages and disadvantages (of a treatment).* [Student #12, male, 14 years]

Another student pointed out a sense of achievement as follows:

> *And later we also searched and found many sources where you can say "yes ok now I have found something useful" and now I can use (the information) to help someone.* [Student #8, female, 13 years]

Results from a brief evaluation questionnaire filled out immediately after the workshop by the students revealed that they were overall satisfied with the workshop and with what they had learned. They described the workshop as "beneficial," "fun," and "cool."

Discussion

Principal Findings

The purpose of this study was to obtain initial insights into the internet-based health information–seeking behavior of children and adolescents aged between 12 and 14 years. We used an

explorative approach to examine the internet-based health information–seeking behavior of the students during an interactive workshop. With this approach, we wanted to gain important insights into a previously sparsely investigated yet crucial subarea of health literacy.

At the beginning of the workshop, all students stated that they actually had limited knowledge about what kind of health information was available on the internet and where to find it. However, they thought that it would be easy to find health-related information on the internet and to evaluate this information.

We found no correlation between this self-assessment of the students and their actual seeking behavior. For example, there was one student who rated internet-based health literacy particularly low but performed a relatively good internet-based search. However, the overall internet-based seeking behavior was rather superficial for nearly all students, regardless of their mostly positive self-view. The first three results presented by the search engine were mainly visited, and the students did not consider whether the information was a paid advertisement. Paid advertisements were mainly by doctors, internet-based magazines, and forums, which were rated as *poor* or *bad* by experts. Among all the visited websites, only two were rated as *good* and the vast majority of websites were rated as either *poor* or *bad*. These results are in line with the findings of previous research [11] showing that youth tend to trust webpages merely because the owner is a doctor or has a similar profession, regardless of the intention (eg, selling products) of the authoring person.

The eHEALS score increased slightly after the workshop, indicating that the students had gained more confidence and competence in the areas of finding and evaluating internet-based health information. After the workshop, the participants were more skeptical about the quality of information on the internet and they were less trusting. This suggests an increased awareness of the students for the large amount of poor-quality health information on the internet.

The results of our study clearly support the findings of previous studies [11,21,44], indicating that students are not used to evaluating internet-based health information critically with objective criteria. This applies to internet users in general, as stated by Sun et al [12], and students in particular [11,21]. To understand the underlying mechanisms of webpages, students need information and training. This training should be given at a young age before they enter the critical phase of adolescence. As youth often develop risky behaviors at this age (eg, risky sexual behavior and smoking or substance abuse), they need to be prepared for critical health issues. Otherwise, the full effects and consequences of such behaviors will become clear in the later stages of life [45].

It is unlikely that the internet's importance as a source of health information will decrease or that nonreliable health-related information will vanish from the internet. There is an urgent need to teach children and adolescents how to properly handle internet-based health information. This could be supported by implementing health literacy in the regular curriculum and developing training programs that foster appropriate and critical internet-based health information seeking. The benefits of such measures would be substantial, not least because increased health literacy can reduce costs for health care systems when individuals rely on high-quality information to make health decisions [46-48].

Limitations

The most important limitation of the study is the small sample size. We cannot assume that the 14 students represent all other Austrian adolescents of similar age. Nevertheless, the results give rise to concerns because despite the low number of students in our evaluation, we expect the results to be reflected in a larger population of students at this age. To prove this assumption, further investigations in this area with larger samples are urgently needed. Additionally, to validate the results, a possible approach could be the application of a computer software that is being newly developed for this purpose.

The study was also limited by the use of the eHEALS, which is based on self-assessment and thus is not an objective instrument for measuring health literacy. The eHEALS scores may be confounded with other competencies, such as reading comprehension and general cognitive abilities. We tried to clarify this aspect, but it remains problematic, because this also applies to the concept of health literacy as a whole. Finally, we noticed a relevant correlation between the eHEALS score before the workshop and the students' general self-efficacy, which might be a confounding factor in measuring health literacy.

Conclusion

The results of this study shed light on the very critical aspects of health literacy and information–seeking behavior. We showed that it is possible to draw the attention of students on the critical aspects of internet search, as well as to slightly improve their competence through a workshop. Although it is very difficult to promote health literacy over a limited time period of 3 days, we could raise awareness of this important topic among students and teachers. Finally, the results highlight that targeted promotion of health literacy, as well as further research is urgently needed, especially among children and adolescents.

Acknowledgments

We would like to express our gratitude to the teachers and students from the school where the workshop took place. We also thank Know-Center Graz for supporting the collection of the data. Further, the workshop was part of the Health-Literacy und Diversity Project, and we would like to thank Gesundheitsfonds Styria for funding this project.

Conflicts of Interest

None declared.

References

1. Henderson EM, Keogh E, Rosser BA, Eccleston C. Searching the internet for help with pain: adolescent search, coping, and medication behaviour. Br J Health Psychol 2013 Feb;18(1):218-232. [doi: 10.1111/bjhp.12005] [Medline: 23126577]
2. Lenhart A. Pew Research Center. 2015. Teens, social media and technology overview 2015: Smartphones facilitate shifts in communication landscape for teens URL: https://www.pewInternet.org/2015/04/09/teens-social-media-technology-2015/ [accessed 2019-04-01]
3. Education Group GmbH. Edugroup. 2017. 5. Oö. Jugend-Medien-Studie 2017 URL: https://www.edugroup.at/innovation/forschung/jugend-medien-studie/detail/5-ooe-jugend-medien-studie-2017.html [accessed 2019-04-01]
4. Eurostat. ec.europa.eu. 2015. Being young in Europe today - digital world URL: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Being_young_in_Europe_today_-_digital_world#Youth_online:_a_way_of_life [accessed 2019-11-01]
5. Wilson TD. Human Information Behavior. Informing Sci J 2000;3:049-056. [doi: 10.28945/576]
6. Zschorlich B, Gechter D, Janßen IM, Swinehart T, Wiegard B, Koch K. Health information on the Internet: Who is searching for what, when and how? Z Evid Fortbild Qual Gesundhwes 2015;109(2):144-152. [doi: 10.1016/j.zefq.2015.03.003] [Medline: 26028452]
7. Bilal D, Gwizdka J. Children's query types and reformulations in Google search. Information Processing & Management 2018 Nov;54(6):1022-1041. [doi: 10.1016/j.ipm.2018.06.008]
8. Kammerer Y, Bohnacker M. Children's web search with Google: the effectiveness of natural language queries. 2012 Jun 13 Presented at: The 11th International Conference on Interaction Design and Children; June 2012; Bremen, Germany p. 184-187. [doi: 10.1145/2307096.2307121]
9. Gwizdka J, Bilal D. Analysis of Children's Queries and Click Behavior on Ranked Results and Their Thought Processes in Google Search. USA: Association for Computing Machinery; 2017 Presented at: The 2017 Conference on Conference Human Information Interaction and Retrieval; 2017; New York, NY p. 377-380. [doi: 10.1145/3020165.3022157]
10. Druin A, Foss E, Hatley L, Golub E, Guha M, Fails J, et al. How children search the internet with keyword interfaces. 2009 Presented at: The 8th International Conference on Interaction Design and Children; 2009; Como, Italy. [doi: 10.1145/1551788.1551804]
11. Subramaniam M, St. Jean B, Taylor NG, Kodama C, Follman R, Casciotti D. Bit by bit: using design-based research to improve the health literacy of adolescents. JMIR Res Protoc 2015 May 29;4(2):e62 [FREE Full text] [doi: 10.2196/resprot.4058] [Medline: 26025101]
12. Sun Y, Zhang Y, Gwizdka J, Trace CB. Consumer Evaluation of the Quality of Online Health Information: Systematic Literature Review of Relevant Criteria and Indicators. J Med Internet Res 2019 May 02;21(5):e12522 [FREE Full text] [doi: 10.2196/12522] [Medline: 31045507]
13. Bromme R, Goldman SR. The Public's Bounded Understanding of Science. Educational Psychologist 2014 Jun 05;49(2):59-69. [doi: 10.1080/00461520.2014.921572]
14. Britt MA, Aglinskas C. Improving Students' Ability to Identify and Use Source Information. Cognition and Instruction 2010 Jun 07;20(4):485-522. [doi: 10.1207/s1532690xci2004_2]
15. Bråten I, Strømsø HI, Britt MA. Trust matters: examining the role of source evaluation in students' construction of meaning withinacross multiple texts. Read Res Q 2011 Nov;44(1):6-28. [doi: 10.1598/rrq.44.1.1]
16. Brand-Gruwel S, Stadtler M. Solving information-based problems: Evaluating sources and information. Learning and Instruction 2011 Apr;21(2):175-179. [doi: 10.1016/j.learninstruc.2010.02.008]
17. Eastin MS, Yang MS, Nathanson AI. Children of the Net: An Empirical Exploration Into the Evaluation of Internet Content. Journal of Broadcasting & Electronic Media 2006 Jun;50(2):211-230. [doi: 10.1207/s15506878jobem5002_3]
18. Braasch JL, Bråten I, Strømsø HI, Anmarkrud Ø, Ferguson LE. Promoting secondary school students' evaluation of source features of multiple documents. Contemporary Educational Psychology 2013 Jul;38(3):180-195. [doi: 10.1016/j.cedpsych.2013.03.003]
19. Walraven A, Brand-Gruwel S, Boshuizen HP. How students evaluate information and sources when searching the World Wide Web for information. Computers & Education 2009 Jan;52(1):234-246. [doi: 10.1016/j.compedu.2008.08.003]
20. Macedo-Rouet M, Braasch JL, Britt MA, Rouet JF. Teaching Fourth and Fifth Graders to Evaluate Information Sources During Text Comprehension. Cognition and Instruction 2013 Apr;31(2):204-226. [doi: 10.1080/07370008.2013.769995]
21. Park E, Kwon M. Health-Related Internet Use by Children and Adolescents: Systematic Review. J Med Internet Res 2018 Apr 03;20(4):e120 [FREE Full text] [doi: 10.2196/jmir.7731] [Medline: 29615385]
22. Couper MP, Singer E, Levin CA, Fowler FJ, Fagerlin A, Zikmund-Fisher BJ. Use of the Internet and ratings of information sources for medical decisions: results from the DECISIONS survey. Med Decis Making 2010;30(5 Suppl):106-114. [doi: 10.1177/0272989X10377661] [Medline: 20881159]

23. Fox S. Pew Research Center. 2011. The Social Life of Health Information, 2011 URL: https://www.pewInternet.org/2011/05/12/the-social-life-of-health-information-2011/ [accessed 2019-04-01]
24. Statistisches Amt der Europäischen Union (Eurostat). data.europa.eu. 2014 Nov. Personen, die das Internet zur Beschaffung von gesundheitsrelevanten Informationen genutzt haben URL: https://data.europa.eu/euodp/de/data/dataset/j2ydOp1Brd6yd7msmmKOsg [accessed 2019-04-01]
25. Wartella E, Rideout V, Montague H, Beaudoin-Ryan L, Lauricella A. Teens, Health and Technology: A National Survey. MaC 2016 Jun 16;4(3):13. [doi: 10.17645/mac.v4i3.515]
26. Keselman A, Arnott Smith C, Murcko AC, Kaufman DR. Evaluating the Quality of Health Information in a Changing Digital Ecosystem. J Med Internet Res 2019 Feb 08;21(2):e11129 [FREE Full text] [doi: 10.2196/11129] [Medline: 30735144]
27. Rice RE. Influences, usage, and outcomes of Internet health information searching: multivariate results from the Pew surveys. Int J Med Inform 2006 Jan;75(1):8-28. [doi: 10.1016/j.ijmedinf.2005.07.032] [Medline: 16125453]
28. Gray NJ, Klein JD, Noyce PR, Sesselberg TS, Cantrill JA. Health information-seeking behaviour in adolescence: the place of the internet. Soc Sci Med 2005 Apr;60(7):1467-1478. [doi: 10.1016/j.socscimed.2004.08.010] [Medline: 15652680]
29. Reid Chassiakos YL, Radesky J, Christakis D, Moreno MA, Cross C. Children and Adolescents and Digital Media. Pediatrics 2016 Nov;138(5):e20162593 [FREE Full text] [doi: 10.1542/peds.2016-2593] [Medline: 27940795]
30. Kuiper E, Volman M, Terwel J. Developing Web literacy in collaborative inquiry activities. Computers & Education 2009 Apr;52(3):668-680. [doi: 10.1016/j.compedu.2008.11.010]
31. Bildungsministerium. bmbwf.gv.at. 2018 Apr. Bildung, Wissenschaft und Forschung URL: https://www.bmbwf.gv.at/Themen/schule/schulpraxis/uek/gesundheit.html [accessed 2019-11-01]
32. Leahy D, Burrows L, McCuaig L, Wright J, Penney D. School Health Education in Changing Times. London: Routledge; Dec 21, 2017.
33. U.S. Department of HealthHuman Services. Healthy People 2010: Understanding and Improving Health. Washington, DC: U.S. Department of HealthHuman Services; 2000.
34. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977 Mar;84(2):191-215. [doi: 10.1037//0033-295x.84.2.191] [Medline: 847061]
35. Bandura A. Perceived Self-Efficacy in Cognitive Development and Functioning. Educational Psychologist 1993 Mar;28(2):117-148. [doi: 10.1207/s15326985ep2802_3]
36. Jerusalem M, Hopf D. Selbstwirksamkeit und Motivationsprozesse in Bildungsinstitutionen. Zeitschrift für Pädagogik 2002;44:54-82. [doi: 10.1163/2214-8647_bnp_e1226050]
37. Weiß RH. Grundintelligenztest Skala 2 - Revision (CFT 20-R). 1st edition. Göttingen: Hogrefe; 2006.
38. Schneider W, Schlagmüller M, Ennemoser M. LGVT 6-12: Lesegeschwindigkeits- und -verständnistest für die Klassen 6-12. Göttingen: Hogrefe; 2007:6.
39. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
40. Jerusalem M, Schwarzer R. Selbstwirksamkeit. In: Skalen zur Befindlichkeit und Persönlichkeit. Berlin: Institut for Psychology, Freie Universität Berlin; 1986:15-28.
41. Jerusalem M, Schwarzer R. Skalen zur Erfassung von Lehrer- und Schülermerkmalen. Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen. Berlin: Freie Universität Berlin; 1999.
42. Pérez A, Potocki A, Stadtler M, Macedo-Rouet M, Paul J, Salmerón L, et al. Fostering teenagers' assessment of information reliability: Effects of a classroom intervention focused on critical source dimensions. Learning and Instruction 2018 Dec;58:53-64. [doi: 10.1016/j.learninstruc.2018.04.006]
43. IBM Corp. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp; 2017.
44. Seiffge-Krenke I. Gesundheit als aktiver Gestaltungsprozess im menschlichen Lebenslauf. In: Oerter R, Montada L, editors. Entwicklungspsychologie, 6th edition. Weinheim: Beltz PVU; 2008:822-839.
45. Gondoli DM. Adolescent development and health. In: Life-span perspectives on health and illness. London: Psychology Press; 1998:159-176.
46. Institute of Medicine Committee on Healthy Literacy. In: Nielsen-Bohlman L, Panzer AM, Kinding DA, editors. Health Literacy: A Prescription To End Confusion. Washington, DC: National Academies Press; 2004.
47. Baker DW, Gazmararian JA, Williams MV, Scott T, Parker RM, Green D, et al. Functional health literacy and the risk of hospital admission among Medicare managed care enrollees. Am J Public Health 2002 Aug;92(8):1278-1283. [doi: 10.2105/ajph.92.8.1278] [Medline: 12144984]
48. Baker DW, Parker RM, Williams MV, Clark WS. Health literacy and the risk of hospital admission. J Gen Intern Med 1998 Dec;13(12):791-798 [FREE Full text] [doi: 10.1046/j.1525-1497.1998.00242.x] [Medline: 9844076]

Abbreviations

eHEALS: electronic health literacy scale

eHealth: electronic health

Edited by G Eysenbach; submitted 17.09.19; peer-reviewed by Y Sun, N Werts, N Taylor, JR Bautista; comments to author 11.11.19; revised version received 05.03.20; accepted 24.03.20; published 26.05.20.

Please cite as:
Maitz E, Maitz K, Sendlhofer G, Wolfsberger C, Mautner S, Kamolz LP, Gasteiger-Klicpera B
Internet-Based Health Information–Seeking Behavior of Students Aged 12 to 14 Years: Mixed Methods Study
J Med Internet Res 2020;22(5):e16281
URL: https://www.jmir.org/2020/5/e16281
doi:10.2196/16281
PMID:32209532

Original Paper

The Role of Psychological Factors in Older Adults' Readiness to Use eHealth Technology: Cross-Sectional Questionnaire Study

Lenka Knapova[1,2,3*], MA; Adam Klocek[1,4*], MA; Steriani Elavsky[1,2,3*], PhD

[1]Institute for Research on Children, Youth and Family, Masaryk University, Brno, Czech Republic

[2]Faculty of Informatics, Masaryk University, Brno, Czech Republic

[3]Faculty of Education, University of Ostrava, Ostrava, Czech Republic

[4]Faculty of Social Studies, Masaryk University, Brno, Czech Republic

[*]all authors contributed equally

Corresponding Author:
Lenka Knapova, MA
Institute for Research on Children, Youth and Family
Masaryk University
Joštova 218/10
Brno, 60200
Czech Republic
Phone: 420 549498272
Email: knapova@mail.muni.cz

Abstract

Background: Information and communication technology (ICT) use among older adults has been on the rise in recent years. However, the predictors and mechanisms behind older adults' acceptance and use of ICT are not clear.

Objective: This study aimed to systematically describe ICT usage among Czech older adults and to evaluate the factors influencing their ICT use and readiness to use digital technology to promote health (eHealth readiness). The primary focus was on psychological factors and the role of persons close to older adults.

Methods: The research utilized cross-sectional survey data from a quota-based sample of Czech older adults (>50 years) and persons close to them further referred to as *close persons* (N=250 dyads). A structural equation modeling framework was used to evaluate relationships between psychological factors, ICT use, and eHealth readiness.

Results: Czech older adults' use of ICT is low with the exception of cell phone usage (cell phone usage by 173/250, 69.2%; other devices used by 50/250, 20.0% of older adults or less). Apart from age (β=−.21; *P*<.001), eHealth readiness was predicted by ICT use (β=.65; *P*<.001). eHealth readiness was also indirectly affected by the need for cognitive closure (NFCC): individuals with a high need for closure perceived more barriers to ICT (β=.23; *P*=.01) and more reported barriers were linked to lower ICT usage (β=−.21; *P*=.001). The expected positive relationships between eHealth readiness of persons close to older adults and ICT use and eHealth readiness of older adults were not significant, but the total effect of eHealth readiness of persons close to older adults on eHealth readiness of older adults was positive and significant (β=.18; *P*=.01), indicating some level of influence of persons close to them on older adults' attitudes and behaviors.

Conclusions: This study provided the first systematic examination of Czech older adults' ICT usage and eHealth readiness. Novel predictors (NFCC and close persons' variables) were evaluated and yielded actionable results. More research is needed to clarify the role of persons close to older adults.

(J Med Internet Res 2020;22(5):e14670) doi:10.2196/14670

KEYWORDS

eHealth; information technology; need for cognitive closure; elderly

Introduction

Background

Older adults (>65 years) are the fastest-growing segment of the population, estimated to account for more than 25% of the total population by 2050 and outnumbering the youngest segment of children under the age of 15 years by 2045 [1,2]. In sharp contrast with the youngsters, older adults are usually portrayed as uninterested in the ever-evolving technological advancements, and thus, lacking familiarity with information and communication technologies (ICTs). However, recent data indicate a sharp increase in technology use among older adults globally, and the Czech Republic has been no exception to this trend. Whereas in 2012, 79.4%, 56.2%, and 17.3% of Czech adults aged 45 to 54 years, 55 to 64 years, and older than 65 years, respectively, used a computer, these numbers have increased to 88.8%, 73.3%, and 32.8% for the respective age groups in 2017 [3]. A similar increase can be seen in the United States, where the percentage of smartphone users older than 65 years rose from 18% in 2013 to 42% in 2017 [4]. Data refined by age and economic activity further support the trend of seniors as the fastest-growing segment of ICT users. The subgroup with the largest increase in the percentage of computer users among Czechs are pensioners, particularly the retired, with only 20.9% of computer users in 2012 compared with 36.1% in 2017 [3].

Considering the position of the Czech Republic within the European Union (EU), rates of ICT usage are slightly below the average for member countries of the EU. For instance, the percentage of households connected to the internet in the Czech Republic is 81.7% as compared with the EU average of 85.4%. The disparities are much more evident in the older segment of the population and for new technological developments such as smartphones and mobile internet. Only 13.1% of Czech adults aged between 55 and 74 years use mobile phones to connect to the internet, which stands in sharp contrast to the average of 34.2% for EU member countries. Within the EU, the Netherlands, Denmark, and Luxembourg rank the highest in terms of older adults using mobile internet, with 61.8%, 61.4%, and 61.1%, respectively. Curiously, the Czech Republic lags behind even when compared with similar countries (by gross domestic product per capita) such as Hungary, Croatia, and Slovakia with 25.2%, 24.2%, and 18.6% of older mobile internet users, respectively [3]. The lower adoption rates of new ICT technologies (eg, mobile internet) among Czech older adults beg the question of how ready Czech older adults are for the deployment of novel ICT approaches in the field of health and medical services, a notion that has been supported by institutions such as the US Department of Health and Human Services or the European Commission and is incorporated within the EU's strategic goals (eg, Europe 2020 strategy) [5]. If left to persist or widen, the digital divide could exacerbate isolation of older adults and may further disadvantage older adults when seeking services that are rapidly being transformed into the digital domain (eg, mobile banking and Web-based reservation systems) or prevent them from adopting newer technologies (eg, assistive technologies) [6,7].

Information and Communication Technology for Health

The use of ICT has been argued to have the potential to positively influence older adults' well-being, decrease loneliness, and increase social support and integration into society [8-10]. This could be especially true with new emerging internet-connected or mobile technologies aimed specifically at improving health and health-related behaviors. Specifically, the term electronic health (eHealth) is used to describe healthy behavior promoting interventions and devices which make use of the internet (eg, computers, personal digital assistants, cell phones, smartphones). Mobile health (mHealth) denotes the use of mobile handheld devices in support of medical and public health practice [11]. The use of technologies such as smartphones, smartphone-based apps, and integrated or connected sensors provides new ways to monitor and improve one's health, healthy lifestyle, and overall well-being. To the point, recent studies support the effectiveness of such mHealth or eHealth interventions for improving health-related behaviors, for example, increased physical activity, sleep, or reduced sedentary behavior [12,13]. For example, Muellman et al [12] conducted a systematic review and found that eHealth interventions promoting physical activity delivered through computer or handheld devices led to increased levels of physical activity in adults aged 55 years and older. Similarly, in a recent review, Elavsky et al [13] concluded that there is evidence supporting the effectiveness of mHealth interventions (defined as treatment programs delivered at least partially by a mobile phone, a smartphone, or a tablet) for increasing physical activity and reducing sedentary behavior in adults aged 50 years and older.

Nevertheless, a vital premise for the effectiveness of eHealth interventions is the initial acceptance of the intervention or program as well as adherence to it over time. Uptake of health-promoting technologies is very low even in carefully conducted research conditions [14], and the continued use of the eHealth technology significantly decreases over time [15]. One factor that can contribute both to the low adherence rates observed in existing interventions as well as the slow uptake of eHealth technologies may be low eHealth readiness. Evaluating how prepared individuals are to adopt eHealth technology or intervention might be exceptionally beneficial when studying older adults whose current adoption and usage of eHealth technologies might be low.

Factors Influencing Information and Communication Technology Acceptance

From a socio-ecological perspective, ICT acceptance among older adults could be best explained by the interaction of factors at multiple levels. At the individual level, ICT use and adoption of new technological developments has been consistently associated with sociodemographic characteristics including age [16-19], education [17-20], income or socioeconomic status [18,21], and marital status [17,18,20]. Vroman et al [17] summarize that nonusers tend to be 75 years or older, have a disability or a chronic health problem, live alone, be single or widowed, and have a lower level of education than ICT users.

Among psychological characteristics, explanations of technology use focus on individual differences in motivations to use technology (eg, technology acceptance model and its variations or hedonic motivation) [22,23], attitudinal factors (eg, perceived ease or usefulness) [24], self-efficacy toward the system [25,26], or anxiety toward the system [26]. In addition, better cognitive abilities, including memory, learning, and concentration, have been linked to higher technology usage [18,27-29]. Dispositional personality characteristics have been acknowledged as potential drivers of ICT adoption and use among older adults as well [17,18]. However, few studies have been conducted so far. One recent study evaluated 17 individual difference predictors of ICT use, indexed with a checklist of 10 different ICTs, including ICTs for health [30]. Need for cognition, defined as an individual's tendency to engage in and enjoy activities that require thinking [31], perceived mastery, and optimism, was found to positively predict ICT use, whereas cynical hostility emerged as a negative predictor [30]. Need for cognition also positively influenced the perceived benefits of ICT use and negatively influenced perceived barriers of ICT use. However, this study was exploratory and included a large number of hypotheses evaluated on one sample without performing any correction for multiple comparisons. Another personality trait which has not been widely studied in the context of ICT use and adoption but might have important implications is the need for cognitive closure (NFCC). NFCC is defined as a desire for a definite answer and an aversion toward ambiguity [32]. Interestingly, NFCC represents both a dimension of stable individual differences and a situationally evocable state [33]. Individuals high in dispositional NFCC feel a desire to quickly reach firm decisions; they are reluctant to have their decisions and views challenged, and they are resistant to information inconsistent with their views [33]. High NFCC individuals were found to be reluctant to change [34] and to be less willing to use innovative technology [35]. Individuals high in NFCC could thus exhibit more negative attitudes toward ICT, including new and innovative technology for improving health, be less prepared to use it, and show lower adoption levels.

Going beyond individual characteristics, other factors in older adults' environments that have been found to affect technology usage of older adults include accessibility, financial support, hardware/software capacity and compatibility, and importantly also support, training, and assistance from others [18,36,37]. Vroman et al [17] add that older adults might be introduced to using ICT for its utility through natural exchanges that occur between family and friends (eg, by receiving hyperlinks via email to view family photos, products, or holiday destinations) and further use ICT as a utility (to search for health-related information, product, and services) and possibly also connect with a virtual community outside one's geographical location. Indeed, previous studies have found that the involvement of family members, especially older adults' children, positively influences ICT adoption [38]. More than 25% of Czech older adults reported that it was their children, grandchildren, or friends who brought them to use the internet [39]. In a similar vein, Vroman et al [17] found that older adults who live alone, are potentially isolated, and lack a social network are the least likely to use ICT.

This Research

The objective of this study was to examine ICT use and eHealth readiness of Czech older adults and to evaluate the influence of psychological factors on older adults' readiness to use eHealth technology, while considering the role of older adults' close persons (eg, children and friends). It was hypothesized that older adults with a high NFCC would perceive more barriers to using ICT, use it less, and exhibit lower eHealth readiness. It was also hypothesized that close persons' eHealth readiness would be positively related to older adults' ICT use and eHealth readiness. Finally, as ICT use has been consistently demonstrated to decline with age and age has been shown to be an important predictor of ICT-related attitudes and behavior, the explanatory models tested considered the effect of age when evaluating the associations among NFCC, perceived barriers, current ICT use, and eHealth readiness.

Methods

Sample and Data Collection

The study was approved by the University Ethical Committee. A total of 250 Czech older adults and the persons close to them (close persons, N=250 dyads) participated in this research. The data were collected between September and November 2017. Participants were recruited and surveyed through a professional marketing and social research agency using stratified quota sampling. The quotas were set based on most recent census data (Czech Statistical Office) to correspond with the underlying population of adults aged 50 years and older based on region (representation of all 14 regions within the Czech republic with a quota based on resident population within each region), gender, age (50-59 years, 60-69 years, 70 years and older), education, and city size (categorized by number of inhabitants). The resulting primary sample of older adults is thus representative of the overall Czech population of adults of 50 years and older in terms of distribution by age, gender, education, region, and city size. Professional interviewers located in various regions were given quota breakdowns and conducted in-person questionnaires with corresponding participants until the quota was met. Each older respondent from the primary sample identified a close person (such as an adult child, a partner, or a friend) who at least occasionally helps them with day-to-day activities (eg, shopping, doctor's visits, household chores, or running errands) with whom they are in contact at least once a week. All close persons were subsequently interviewed either in person or by telephone. The data from the primary sample of older adults were collected through standardized, structured face-to-face interviews (approximately 45 min in length) in the households of respondents with the help of a tablet and a questionnaire software. The data were collected at one time point and are cross-sectional.

Measures

Demographics

Basic demographic information was collected (ie, age, gender, education, income, residence).

Information and Communication Technology Use

Participants provided information about the following devices: computer, laptop, cell phone, smartphone, and tablet. Participants were first asked to indicate the devices they own and use and subsequently report further details such as the daily usage time (hours/day) and length of use. This information was aggregated to form two measures of ICT use: *number of devices used* and *total usage time*. The *number of devices used* was calculated as a sum of all the devices participants reported to use (not own). The *total usage time* of the ICT devices per day was calculated as the sum of usage times per day provided for the specific devices. When a participant did not report usage of a device (and thus was not asked about the details of its use), the usage time was coded as 0. Responses for total reported usage time exceeding 24 hours per day (3 older adults, 12 close persons) were recoded to a maximum value of 16 hours per day to account for typical sleep duration and basic needs.

In addition, the use of the internet as a specific ICT-related technology was assessed. Participants were asked if they use the internet. Internet users were asked for further information about their *internet use,* such as the daily usage time (hours/day), length of use, and frequency of use. For nonusers of the internet, the usage time was coded as 0.

Perceived Barriers to Information and Communication Technology Use

Participants chose from the following list of barriers generated from previous studies on the topic: *Not enough knowledge*; *Technologies evolve too fast; I have trouble with vision and psychomotorics; I am afraid of making a mistake; I do not have enough support; Too expensive, I cannot afford it; I don't care about internet; Something else.* Participants could indicate multiple barriers with the total number of perceived barriers ranging from 0 to 7.

Electronic Health Readiness

Both older adults and their close persons completed the eHealth Readiness scale by Bhalla et al [40]. The scale was translated to Czech using the method of double-back translation, with emphasis on systematic equivalence [41]. Inconsistencies in item wording were resolved by group discussion of study authors, which included senior researchers on the topic and bilingual speakers. The items were scored on a 6-point scale ranging from 1 (*completely disagree*) to 6 (*completely agree*). The total score was calculated as the sum of all the items. Bhalla et al [40] reported a Cronbach alpha of .81 and .83 for validation samples. Internal consistency in this research was higher: Cronbach alpha was .91 and .87 for the older adults and their close persons, respectively.

Need for Cognitive Closure

NFCC was measured with the Czech version of the 15-item NFCC Scale [42,43]. The items were scored on a 5-point scale ranging from 1 (*completely disagree*) to 5 (*completelyagree*). The total score was calculated as the arithmetical mean of all the items. Širůček et al [43] reported a Cronbach alpha of .84. The Cronbach alpha in this research was .87.

Data Analysis

Statistical analyses were performed using R (R Foundation for Statistical Computing), version 3.4.2 [44], packages psych (version 1.8.3.3), lavaan (version 0.5-23.1097), semPlot (version 1.1), and semTools (version 0.4-14).

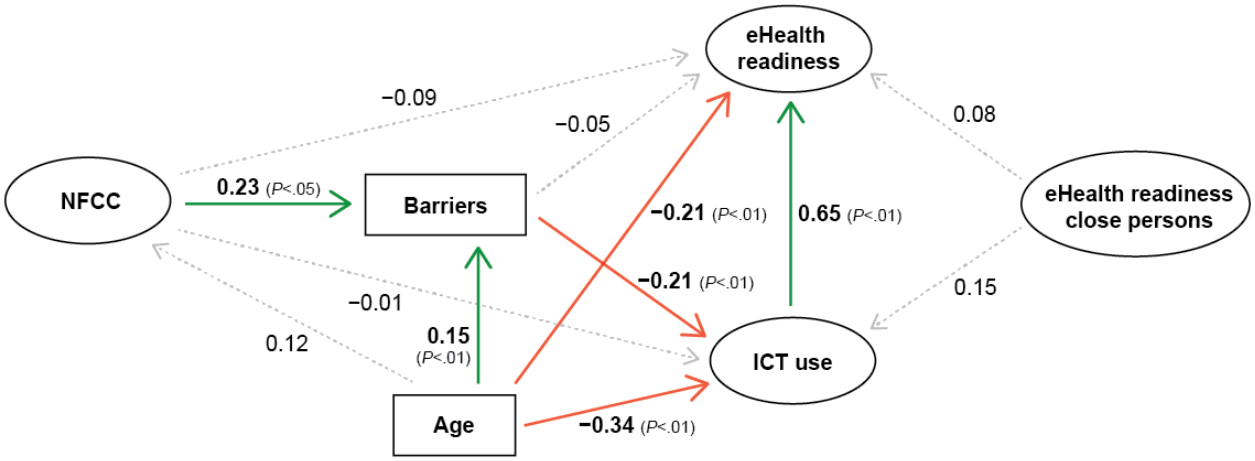
The proposed relationships were tested using structural equation modeling (SEM). The structural model was estimated using the maximum likelihood with robust SEs test statistics estimator. The model fit was evaluated using standard measures of model fit: the standardized root mean square residual (SRMR), which should be less than or approximately 0.08 [45,46]; the root mean square error of approximation (RMSEA), with values below 0.05 indicating a close fit and values below 0.08 being less indicative of a good fit [47,48]; the comparative fit index (CFI), which should be higher than 0.90 [49]; and the Tucker-Lewis index (TLI) with recommended values greater than 0.90 [50] or 0.95 [46].

Model Testing

First, we defined three latent variables. The latent variable of ICT use was specified as the combination of three manifest variables: number of used devices, total usage time of the devices, and internet usage time. The measurement model for eHealth readiness was a simple 1-factor model where all items were loaded on a single latent variable. eHealth readiness item 2 was excluded from the scale in both samples as it exhibited high residual correlations with other variables and worsened the overall fit of both the measurement model and the resulting structural model. Moreover, upon closer examination, the item (*I feel that my previous experience with online technologies is important to my success with using a lifestyle intervention*) intertwines previous experience with opinions about its importance, and the item wording makes it difficult to decide if an individual who has previous experience with online technologies but does not think this experience is important to one's success should be labeled as more or less ready to use eHealth technology. For NFCC, a 1-factor model was defined where all NFCC items loaded on a single latent variable and error correlations between items from the same facet were allowed to account for their common facet source (ie, residual correlations were allowed for triads of items 1-3, 4-6, 7-9, 10-12, and 13-15).

On the basis of research questions and the hypothesized relationships, the structural model portrayed in Figure 1 was tested. With respect to the association between older adults' and close persons' eHealth readiness, the residual covariance between the items from the eHealth readiness scale (specifically item 1) was allowed based on residual correlation matrices, modification suggestions, as well as to account for the dyadic nature of the data and possible shared environment.

Figure 1. A simplified version of the tested model with path estimates. Ellipses indicate latent variables; rectangles indicate manifest variables; full colored arrows indicate significant relationships (green: positive; red: negative) whereas dashed arrows indicate nonsignificant relationships. eHealth: electronic health; ICT: information and communication technology; NFCC: need for cognitive closure.



Results

Sample Characteristics

The mean age in the older adult sample was 66.14 (SD 9.47) years; 55.2% (138/250) were women. The majority of the older adults were retired (188/250, 75.2%) and completed high school education (171/250, 68.4%). In the sample of close persons, the mean age was 46.30 (SD 13.51) years; 70.4% (176/250) were women. Regarding the relationship toward the older adult, 54.8% (137/250) of close persons were children, 18.0% (45/250) were other relatives, and 15.2% (38/250) were partners, the rest accounting for friends, acquaintances, or professional caretakers. Detailed demographic characteristics of both samples can be found in Table 1.

The most widely used device among older adults was the cell phone, with 69.2% (173/250) of older adults using it on a daily basis. All the other devices were used by less than a fourth of the participants. Details on the mean daily usage (hours/day) of the ICT devices and internet among daily users as well as the total usage time and the total number of devices among all older adults can be found in Table 2. The table also provides details on the number of perceived barriers, NFCC, and eHealth readiness (for the older adults and their close persons as well).

Differences based on gender were examined and are portrayed in Table 2. In total, men used more ICT devices than women, and this difference was significant (t_{248}=2.12 [2-talied *t* test]; *P*=.04; Cohen *d*=0.27). No other differences based on gender were significant.

As for the aggregated ICT usage variables, there were significant differences based on education in the total usage time ($F_{3,246}$=3.23; *P*=.02), the number of used devices ($F_{3,246}$=5.40; *P*=.001), as well as the number of perceived barriers ($F_{3,246}$=3.72; *P*=.01). For usage time, college-educated participants reported higher time than participants with primary education, but the difference was not significant (mean difference 2.44; *P*=.05). Participants with primary education used fewer devices than participants with a high-school diploma (mean difference 0.43; *P*=.008) and to a lesser degree than participants with college education (mean difference 0.56; *P*=.005). Regarding the number of perceived barriers, the only significant group difference was between participants with primary education who perceived more barriers than college-educated participants (mean difference 0.87; *P*=.02). The most commonly perceived barriers by the older adults related to technologies evolving too fast (indicated by 96/250, 38.4% of older adults), fear of making a mistake (74/250, 29.6%), not being interested in internet technology (55/250, 22.0%), technologies being too expensive (48/250, 19.2%), and not having enough knowledge (47/250, 18.8%).

When comparing older adults with their close persons, close persons used on average more ICT devices (mean 2.03, SD 0.91; t_{249}=−11.91; *P*<.001; Cohen *d*=0.958), used them for more hours per day (mean 5.32, SD 4.32; t_{249}=−10.49; *P*<.001; Cohen *d*=0.76), and perceived fewer barriers to using ICT technology (mean 0.62, SD 0.95; t_{249}=9.73; *P*<.001; Cohen *d*=0.79). Nevertheless, the sample of close persons was also significantly younger than the primary sample of older adults (t_{249}= 22.55; *P*<.001; Cohen *d*=1.70).

Table 1. Demographic characteristics of older adults and their close persons.

Characteristic	Older adults	Close persons
Age (years)		
Mean (SD)	66.14 (9.47)	46.30 (13.51)
Range	50-93	21-90
Gender, n (%)		
Females	138 (55.2)	176 (70.4)
Males	112 (44.8)	74 (29.6)
Education, n (%)		
Elementary	55 (22.0)	10 (4.0)
Secondary school (no diploma)	101 (40.4)	93 (37.2)
Secondary school (diploma)	70 (28.0)	108 (43.2)
University	24 (9.6)	39 (15.6)
Marital status, n (%)		
Married	88 (35.2)	133 (53.2)
Divorced	60 (24.0)	47 (18.8)
Widowed	89 (35.6)	9 (3.6)
Single	7 (2.8)	44 (17.6)
Living with a partner	6 (2.4)	17 (6.8)
Occupation, n (%)		
Employed	51 (20.4)	172 (68.8)
Retired	188 (75.2)	44 (17.6)
Unemployed	11 (4.4)	28 (11.2)
Students	N/A[a]	6 (2.4)

[a]N/A: not applicable.

Table 2. Descriptive characteristics for the information and communication technology use variables, electronic health readiness, and need for cognitive closure.

Characteristic	Proportion of daily users (N=250), n (%)	Entire sample		Females		Males	
		Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
Cell phone usage (hours/day)	173 (69.2)	1.89 (3.31)	0.1-24.0	2.21 (4.22)	0.1-24.0	1.54 (1.82)	0.1-10.0
Smartphone usage (hours/day)	19 (7.6)	2.05 (1.92)	0.5-8.0	1.95 (1.61)	0.5-6.0	2.17 (2.32)	0.5-8.0
Personal computer usage (hours/day)	50 (20.0)	3.16 (2.52)	1.0-10.0	3.00 (2.49)	1.0-8.0	3.29 (2.60)	1.0-10.0
Laptop usage (hours/day)	39 (15.6)	2.64 (1.70)	0.1-7.0	2.49 (1.55)	0.1-6.0	2.81 (1.89)	0.5-7.0
Tablet usage (hours/day)	7 (2.8)	1.71 (0.76)	1.0-3.0	1.33 (0.58)	1.0-2.0	2.00 (0.82)	1.0-3.0
Internet usage (hours/day)	99 (39.6)	2.24 (1.98)	0.1-10.0	2.23 (1.92)	0.1-10.0	2.47 (2.02)	0.2-10.0
Total usage time (hours/day)	N/A[a]	2.43 (3.26)	0-16	2.26 (3.25)	0-16	2.64 (3.28)	0-14
Number of devices	N/A	1.24 (0.73)	0-4	1.15 (0.71)	0-4	1.35 (0.74)	0-4
Number of barriers	N/A	1.47 (1.20)	0-6	1.59 (1.25)	0-6	1.32 (1.13)	0-5
Need for cognitive closure	N/A	3.56 (0.56)	1.80-5.00	3.60 (0.54)	2.13-4.80	3.51 (0.59)	1.80-5.00
eHealth[b] readiness	N/A	15.56 (7.55)	6-33	15.03 (7.49)	6-33	16.21 (7.60)	6-32
eHealth readiness close persons	N/A	22.76 (7.28)	6-36	23.07 (7.07)	6-36	22.38 (7.55)	6-35

[a]N/A: not applicable.

[b]eHealth: electronic health.

Predicting Information and Communication Technology Use and eHealth Readiness

The evaluated structural model with standardized estimates of the regression paths is depicted in Figure 1. The model fit was good, χ^2_{438}=804.1; the χ^2 to degree of freedom ratio was 1.84; CFI=0.911; TLI=0.899; SRMR=0.064; RMSEA=0.060; 95% CI (0.053-0.066), considering the complexity of the model, the sample size, and the initial fit of individual latent variables.

Estimates for the direct effects are displayed in Table 3. The results showed that the older an adult, the more barriers to using technology he/she perceived (β=.15; *P*=.008) and the lower his/her ICT use (β=−.34; *P*<.001) and eHealth readiness were (β=−.21; *P*<.001). Individuals with a high NFCC also perceived more barriers (β=.23; *P*=.01). Older adults who reported more barriers in fact used ICT less (β=−.21; *P*=.001). The relationship between ICT use and eHealth readiness of older adults was positive, meaning that individuals who use ICT more are also more prepared and willing to accept eHealth technology (β=.65; *P*<.001). None of the other direct effects were significant.

Indirect effect of NFCC on ICT use through perceived barriers was negative and significant (β=−.049; *P*=.04), meaning that a higher NFCC was related to more perceived barriers. Indirect effect of NFCC on eHealth readiness of older adults through perceived barriers and ICT use was negative and significant as well (β=−.031; *P*=.04). The expected direct effects of eHealth readiness of close persons on ICT use and eHealth readiness of older adults were not significant, but the total effect of eHealth readiness of close persons on eHealth readiness of older adults was positive and significant (β=.18; *P*=.02).

The evaluated model explained 63.7% of the variance of older adults' eHealth readiness, 21.2% of variance of ICT use, and 8.5% of variance of the number of perceived barriers.

Table 3. Estimates of direct effects.

Regression path	Estimate B[a]	SE	*P* value	95% CI	Standardized estimate β
NFCC[b] –> eHealth[c] Readiness	−0.187	0.117	.11	−0.415 to 0.042	−.091
NFCC –> Barriers	0.750	0.301	.01	0.160 to 1.339	.232
NFCC –> ICT[d] use	−0.055	0.574	.92	−1.180 to 1.070	−.008
Barriers –> eHealth Readiness	−0.034	0.028	.23	−0.090 to 0.022	−.053
Barriers –> ICT use	−0.429	0.133	.001	0.689 to −0.169	−.210
ICT use –> eHealth Readiness	0.201	0.034	<.001	0.134 to 0.268	.647
eHealth Readiness Close Persons –> eHealth Readiness	0.118	0.075	.12	−0.028 to 0.264	.080
eHealth Readiness Close Persons –> ICT use	0.730	0.432	.09	−0.117 to 1.577	.154
Age –> eHealth Readiness	−0.017	0.004	<.001	−0.025 to −0.008	−.207
Age –> Barriers	0.019	0.007	.008	0.005 to 0.034	.151
Age –> ICT use	−0.089	0.017	<.001	−0.122 to −0.056	−.342
Age –> NFCC	0.005	0.004	.24	−0.003 to 0.012	.117

[a]Nonstandardized estimate of direct effects (as compared with the standardized estimate β).

[b]NFCC: need for cognitive closure.

[c]eHealth: electronic health.

[d]ICT: information and communication technology.

Discussion

Principal Findings

This study evaluated the predictors of ICT use and readiness to use eHealth technology by older adults. The study considered the role of perceived barriers, the role of close persons' eHealth readiness, and it was the first study to evaluate the influence of NFCC on ICT use and eHealth readiness of older adults. Although unable to definitively establish the direction of the studied relationships because of the cross-sectional nature of the data, the SEM analysis showed that apart from age, eHealth readiness was predicted by ICT use. Older adults who used ICT more in general were more ready to use technology for supporting health and healthy behaviors. Nevertheless, a reverse (higher eHealth readiness predicts more ICT use) or bidirectional relationship could also exist, but these were not specified and tested in the current model. Although the NFCC did not directly impact ICT use or eHealth readiness, NFCC exerted influence on ICT use and eHealth readiness indirectly through the number of perceived barriers. Individuals high in NFCC perceived more barriers to ICT. The number of barriers was, in turn, negatively related to their overall ICT use. Interestingly, there was evidence that significant others might influence older adults' eHealth readiness, although the mechanisms remain unclear (the direct effect of close persons' eHealth readiness on ICT use and eHealth readiness of older adults was not significant, but the total effect of close persons' eHealth readiness on the older adults' eHealth readiness was significant).

This study supported previous research [16-19] by finding that age plays an important role in ICT use and adoption. Older adults used ICT devices less, were less prepared to use it for monitoring and improving health, and perceived more barriers to ICT use than their younger counterparts. Similarly, in line with previous research [17-20], ICT usage was related to the level of education. Differences were found especially between older adults with primary education who used ICT to a lesser degree and perceived more barriers than college-educated older adults.

A novel predictor yielding interesting results was the NFCC. This research built upon the pioneering studies by Chernikova et al [35], which found NFCC to be related to intentions to use technological innovations and support the notion that NFCC may in fact play a role in technology adoption and use. The indirect effect of NFCC on ICT use and eHealth readiness through the number of perceived barriers (and in case of eHealth readiness through ICT use as well) suggests that individuals high in NFCC may perceive more barriers to ICT adoption and use, resulting in lower ICT use and eHealth readiness, but as a function of their reluctance to change and preference for tradition and security rather than the actual inability to use or learn to use ICT [34,51]. This notion should be evaluated in future research and should include a measure of perceived barriers by older adults as well as, for instance, a close person's evaluation of the older adult's actual barriers to technology adoption and use. Interestingly, NFCC indirectly predicted ICT use and eHealth readiness even when the effects of age were accounted for in the model, suggesting that this dispositional characteristic may, in part, help explain some of the age-related decline observed in ICT use.

The obtained results on the role of NFCC and perceived barriers on ICT use and readiness suggest interesting actionable implications. Researchers carrying out ICT use promoting interventions and programs might want to measure the level of NFCC of the participants in their programs. NFCC may then serve as a potential tailoring variable in the design of programs promoting ICT use, wherein depending on the individual's level

of NFCC, intervention components explicitly focusing on reducing the number of perceived barriers could be incorporated into the intervention. The respective strategies to mitigate barriers would depend on the specific barriers but could range from providing training with ICT devices to financial support to presenting older adults with available technology and how it can enhance one's life. Effectively reducing perceived barriers (especially in high NFCC individuals) may then positively impact ICT acceptance and use. Moreover, NFCC has also been shown to be a situationally evocable state. This research suggests that when presenting a new ICT or eHealth device, application, or intervention to its potential users, it may be beneficial to lower their NFCC before and during the description of the device or intervention. This could be done, for example, by providing potential users with sufficient time when making decisions, by stressing the importance of forming an accurate judgment, or by aiding in the process of finding additional information on the matter before forming a judgement [33,52]. Experimental research testing these propositions is needed.

As for the role of significant others on older adults' adoption and use of technology, which has been proposed by several theoretical approaches [17,53], the results are inconclusive at this point, and further research is needed to clarify this relationship. On the basis of these results, close persons might influence older adults' eHealth readiness, but the mechanisms are not clear. The amount, level, and specifics of support of close persons to older adults regarding ICT and specifically eHealth technologies were not explicitly assessed in the present research. Further research could focus on evaluating the specific mechanisms of close persons' influence on older adults' ICT acceptance and use. A better understanding of the role that close persons play in the adoption of ICT technologies by older adults would be particularly useful when designing ICT use promotion interventions for older adults. Close persons could help facilitate ICT adoption by the provision of specific types of support or through other mechanisms yet to be identified.

Limitations and Further Research

Considering the limited amount of research that has been conducted specifically on the relationship between NFCC and technology, this study should be regarded as an exploratory study, and the model should be confirmed and cross-validated on other samples to increase the validity of the results. A limitation of the present research is also the correlational nature of data, which does not allow for firm conclusions about the causality and direction of the relationships. Although the proposed relationships were theoretically construed and the proposed direction of causality seems theoretically plausible, further research should validate the findings in different older adult samples.

The sample size was rather low, considering the complexity of the evaluated models. It is also possible that an equally well-fitting structural model would result from a specification of different relationships. This further underscores the need for cross-validation and empirical testing of competing models in independent samples of older adults. Ideally, this research should involve studies with prospective, longitudinal, or experimental designs to allow for more definitive conclusions regarding the causality and time ordering of relationships under study.

Finally, little research has been conducted on the topic of ICT acceptance and use by Czech older adults—to our knowledge, we present the first systematic examination of ICT use and eHealth readiness in Czech older adults. Considering the rather low ICT usage and eHealth readiness in the current sample, it would be interesting to repeat the survey in a few years' time to evaluate the changes. Similarly, comparing specific subsamples of the older adult population such as older adults who are physically active, older adults visiting university classes, and people living in homes for the elderly could yield information about further factors influencing ICT acceptance and use, helping shape ICT use promotion efforts and policies related to ICT adoption across the population spectrum of older adults.

Conclusions

This is the first study to systematically evaluate Czech older adults' readiness to use technology for improving health, their related ICT use, and possible predictors, including NFCC and the role of significant others. Our results provide new insights into the predictors of older adults' readiness to use eHealth technology, especially with respect to NFCC. eHealth readiness was found to be affected directly by age and actual ICT use and indirectly by the NFCC and the number of perceived barriers toward using technology. These results are directly applicable for researchers or organizations carrying out interventions promoting the use of eHealth devices and applications.

Future researchers are encouraged to validate the findings in various older adult samples and further clarify the role of older adults' close persons. Additional studies, including prospective or experimental studies, are required to support the presented findings.

Acknowledgments

These results are part of the project that has received funding from the EU's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie, and it is cofinanced by the South Moravian Region under grant agreement No. 665860. This material reflects only the author's attitudes, and the EU is not responsible for any possible use of the information contained in such material.

Conflicts of Interest

None declared.

References

1. United Nations. World Population Ageing 2007. New York: United Nations; 2007.
2. United Nations. World Population Prospects: 2015 Revision. New York: United Nations; 2015.
3. Český statistický úřad | ČSÚ.: Cesky statisticky urad; 2017. Use of ICT in Households and Individuals - 2017. Article in Czech. Využívání informačních a komunikačních technologií v domácnostech a mezi jednotlivci - 2017 URL: https://www.czso.cz/csu/czso/vyuzivani-informacnich-a-komunikacnich-technologii-v-domacnostech-a-mezi-jednotlivci-2017 [accessed 2020-03-23] [WebCite Cache ID 78CNVtyXV]
4. Anderson M, Perrin A. Pew Research Center. Washington, DC: Pew Research Center; 2017 May 17. Tech Adoption Climbs Among Older Adults URL: https://www.pewresearch.org/internet/2017/05/17/tech-adoption-climbs-among-older-adults/ [accessed 2020-03-23]
5. European Commission. Brussels: European Commission; 2010 Mar 3. Communication from the Commission. Europe 2020: A Strategy for Smart, Sustainable and Inclusive Growth URL: https://ec.europa.eu/eu2020/pdf/COMPLET%20EN%20BARROSO%20%20%20007%20-%20Europe%202020%20-%20EN%20version.pdf [accessed 2020-03-23]
6. Chesley N, Johnson BE. Information and communication technology use and social connectedness over the life course. Soc Compass 2014 Jun 19;8(6):589-602. [doi: 10.1111/soc4.12170]
7. Damodaran L, Gilbertson T, Olphert W, Sandhu J. Digital inclusion - the vision, the challenges and the way forward. Int J Adv Internet Technol 2015;8(3-4):78-92 [FREE Full text]
8. Backonja U, Hall AK, Thielke S. Older adults' current and potential uses of information technologies in a changing world: a theoretical perspective. Int J Aging Hum Dev 2014 Dec;80(1):41-63 [FREE Full text] [doi: 10.1177/0091415015591109] [Medline: 26215298]
9. Independent Age. Older People, Technology and Community: The Potential of Technology to Help Older People Renew or Develop Social Contacts and to Actively Engage in Their Communities. London: Independent Age; 2010.
10. Czaja SJ, Boot WR, Charness N, Rogers WA, Sharit J. Improving social support for older adults through technology: findings from the prism randomized controlled trial. Gerontologist 2018 May 8;58(3):467-477 [FREE Full text] [doi: 10.1093/geront/gnw249] [Medline: 28201730]
11. World Health Organization. 2016. International Classification of Health Interventions (ICHI) URL: https://www.who.int/classifications/ichi/en/ [accessed 2020-03-23] [WebCite Cache ID 78CPF0d9q]
12. Muellmann S, Forberger S, Möllers T, Bröring E, Zeeb H, Pischke CR. Effectiveness of eHealth interventions for the promotion of physical activity in older adults: A systematic review. Prev Med 2018 Mar;108:93-110. [doi: 10.1016/j.ypmed.2017.12.026] [Medline: 29289643]
13. Elavsky S, Knapova L, Klocek A, Smahel D. Mobile health interventions for physical activity, sedentary behavior, and sleep in adults aged 50 years and older: a systematic literature review. J Aging Phys Act 2019 Aug 1;27(4):565-593. [doi: 10.1123/japa.2017-0410] [Medline: 30507266]
14. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec 5;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
15. Mattila E, Orsama A, Ahtinen A, Hopsu L, Leino T, Korhonen I. Personal health technologies in employee health promotion: usage activity, usefulness, and health-related outcomes in a 1-year randomized controlled trial. JMIR Mhealth Uhealth 2013 Jul 29;1(2):e16 [FREE Full text] [doi: 10.2196/mhealth.2557] [Medline: 25098385]
16. Magsamen-Conrad K, Upadhyaya S, Joa CY, Dowd J. Bridging the Divide: Using UTAUT to predict multigenerational tablet adoption practices. Comput Human Behav 2015 Sep 1;50:186-196 [FREE Full text] [doi: 10.1016/j.chb.2015.03.032] [Medline: 25937699]
17. Vroman KG, Arthanat S, Lysack C. 'Who over 65 is online?' Older adults' dispositions toward information communication technology. Comp Hum Behav 2015 Feb;43:156-166. [doi: 10.1016/j.chb.2014.10.018]
18. Wagner N, Hassanein K, Head M. Computer use by older adults: a multi-disciplinary review. Comput Hum Behav 2010 Sep;26(5):870-882. [doi: 10.1016/j.chb.2010.03.029]
19. Olson KE, O'Brien MA, Rogers WA, Charness N. Diffusion of technology: frequency of use for younger and older adults. Ageing Int 2011 Mar;36(1):123-145 [FREE Full text] [doi: 10.1007/s12126-010-9077-9] [Medline: 22685360]
20. Selwyn N, Gorard S, Furlong J, Madden L. Older adults' use of information and communications technology in everyday life. Ageing Soc 2003 Sep 2;23(5):561-582. [doi: 10.1017/S0144686X03001302]
21. Vorrink SN, Antonietti AM, Kort HS, Troosters T, Zanen P, Lammers JJ. Technology use by older adults in the Netherlands and its associations with demographics and health outcomes. Assist Technol 2017;29(4):188-196. [doi: 10.1080/10400435.2016.1219885] [Medline: 27548101]
22. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. Manag Inf Syst Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
23. Macedo IM. Predicting the acceptance and use of information and communication technology by older adults: an empirical examination of the revised UTAUT2. Comput Hum Behav 2017 Oct;75:935-948. [doi: 10.1016/j.chb.2017.06.013]
24. Marangunić N, Granić A. Technology acceptance model: a literature review from 1986 to 2013. Univ Access Inf Soc 2015;14(1):81-95. [doi: 10.1007/s10209-014-0348-1]

25. Taylor S, Todd P. Decomposition and crossover effects in the theory of planned behavior: A study of consumer adoption intentions. Int J Res Mark 1995 Jul;12(2):137-155. [doi: 10.1016/0167-8116(94)00019-K]
26. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. Manag Inf Syst Q 2003;27(3):425-478. [doi: 10.2307/30036540]
27. Czaja SJ, Charness N, Fisk AD, Hertzog C, Nair SN, Rogers WA, et al. Factors predicting the use of technology: findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). Psychol Aging 2006 Jun;21(2):333-352 [FREE Full text] [doi: 10.1037/0882-7974.21.2.333] [Medline: 16768579]
28. Mitzner T, Savla J, Boot W, Sharit J, Charness N, Czaja S, et al. Technology adoption by older adults: findings from the PRISM Trial. Gerontologist 2019 Jan 9;59(1):34-44 [FREE Full text] [doi: 10.1093/geront/gny113] [Medline: 30265294]
29. Eastman JK, Iyer R. The elderly's uses and attitudes towards the internet. J Consum Mark 2004 May;21(3):208-220. [doi: 10.1108/07363760410534759]
30. Chopik WJ, Rikard R, Cotten SR. Individual difference predictors of ICT use in older adulthood: A study of 17 candidate characteristics. Comput Hum Behav 2017 Nov;76:526-533. [doi: 10.1016/j.chb.2017.08.014]
31. Cacioppo JT, Petty RE. The need for cognition. J Pers Soc Psychol 1982 Jan;42(1):116-131. [doi: 10.1037/0022-3514.42.1.116]
32. Webster DM, Kruglanski AW. Individual differences in need for cognitive closure. J Pers Soc Psychol 1994;67(6):1049-1062. [doi: 10.1037/0022-3514.67.6.1049]
33. Roets A, Kruglanski A, Kossowska M, Pierro A, Hong Y. The motivated gatekeeper of our minds: new directions in need for closure theory and research. Adv Exp Soc Psychol 2015;52:221-283. [doi: 10.1016/bs.aesp.2015.01.001]
34. Kruglanski AW, Pierro A, Higgins ET, Capozza D. 'On the move' or 'Staying put': locomotion, need for closure, and reactions to organizational change. J Appl Soc Pyschol 2007 Jun;37(6):1305-1340. [doi: 10.1111/j.1559-1816.2007.00214.x]
35. Chernikova M, Kruglanski A, Giovannini D, Vezzali L, Su J. Need for closure and reactions to innovation. J Appl Soc Psychol 2017 Jun 23;47(9):473-481. [doi: 10.1111/jasp.12451]
36. Pan S, Jordan-Marsh M. Internet use intention and adoption among Chinese older adults: from the expanded technology acceptance model perspective. Comput Hum Behav 2010 Sep;26(5):1111-1119. [doi: 10.1016/j.chb.2010.03.015]
37. Ryu M, Kim S, Lee E. Understanding the factors affecting online elderly user's participation in video UCC services. Comput Hum Behav 2009 May;25(3):619-632. [doi: 10.1016/j.chb.2008.08.013]
38. Fausset C, Harley L, Farmer S, Fain B. Older Adults' Perceptions and Use of Technology: A Novel Approach. In: Proceedings of the International Conference on Universal Access in Human-Computer Interaction. Proceedings of the 7th International Conference Universal Access in Human-Computer Interaction: User and Context Diversity; 2013 Presented at: UAHCI'13; 21-26 July 2013; Las Vegas p. 51-58. [doi: 10.1007/978-3-642-39191-0_6]
39. Kopecky K, Szotkowski R, Kozisek M, Kasackova J. Projekt E-bezpečí. 2018. Old Men on the Net (Research Report). Article in Croatian. Starci na netu (výzkumná zpráva) URL: https://www.e-bezpeci.cz/index.php/ke-stazeni/vyzkumne-zpravy/102-starci-na-netu-2017-2018/file [accessed 2020-03-23] [WebCite Cache ID 78CWLKXp7]
40. Bhalla A, Durham RL, Al-Tabaa N, Yeager C. The development and initial psychometric validation of the eHealth readiness scale. Comput Hum Behav 2016 Dec;65:460-467. [doi: 10.1016/j.chb.2016.09.015]
41. Maneesriwongul W, Dixon JK. Instrument translation process: a methods review. J Adv Nurs 2004 Oct;48(2):175-186. [doi: 10.1111/j.1365-2648.2004.03185.x] [Medline: 15369498]
42. Roets A, van Hiel A. Item selection and validation of a brief, 15-item version of the Need for Closure Scale. Pers Individ Dif 2011 Jan;50(1):90-94. [doi: 10.1016/j.paid.2010.09.004]
43. Sirucek J, Tapal A, Linhartova P. The need for cognition: A study of the psychometric characteristics of the abbreviated Czech version The Scale of the Need for Cognition. Article in Czech. Potřeba poznávání: Studie psychometrických charakteristik zkrácené české verze Škály potřeby poznávání. Czechoslovak Psychol 2014;58(1):52-61 [FREE Full text]
44. R Core Team. The R Project for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2017. URL: https://www.r-project.org/ [accessed 2020-03-23]
45. Bollen KA. Structural Equations With Latent Variables. New York: Wiley; 1989.
46. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equ Modeling 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
47. Browne MW, Cudeck R. Alternative ways of assessing model fit. In: Bollen KA, Long JS, editors. Testing Structural Equation Models. Newbury Park: Sage; 1993:136-162.
48. MacCallum RC, Browne MW, Sugawara HM. Power analysis and determination of sample size for covariance structure modeling. Psychol Methods 1996;1(2):130-149. [doi: 10.1037/1082-989X.1.2.130]
49. Marsh HW, Hau K, Wen Z. In search of golden rules: comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. Struct Equ Modeling 2004 Jul;11(3):320-341. [doi: 10.1207/s15328007sem1103_2]
50. Bentler PM, Bonett DG. Significance tests and goodness of fit in the analysis of covariance structures. Psychol Bull 1980;88(3):588-606. [doi: 10.1037/0033-2909.88.3.588]
51. Calogero RM, Bardi A, Sutton RM. A need basis for values: associations between the need for cognitive closure and value priorities. Pers Individ Dif 2009 Jan;46(2):154-159. [doi: 10.1016/j.paid.2008.09.019]

52. Kruglanski AW, Webster DM. Motivated closing of the mind: 'seizing' and 'freezing'. Psychol Rev 1996 Apr;103(2):263-283. [doi: 10.1037/0033-295x.103.2.263] [Medline: 8637961]
53. Bronfenbrenner U. The Ecology Of Human Development: Experiments By Nature And Design. Cambridge: Harvard University Press; 1979.

Abbreviations

CFI: comparative fit index
eHealth: electronic health
EU: European Union
ICT: information and communication technology
mHealth: mobile health
NFCC: need for cognitive closure
RMSEA: root mean square error of approximation
SEM: structural equation modeling
SRMR: standardized root mean square residual
TLI: Tucker-Lewis Index

Edited by G Eysenbach; submitted 09.05.19; peer-reviewed by SJ Czaja, P Santana-Mancilla, B Renn, R Hervás; comments to author 03.10.19; revised version received 17.12.19; accepted 09.02.20; published 28.05.20.

Please cite as:
Knapova L, Klocek A, Elavsky S
The Role of Psychological Factors in Older Adults' Readiness to Use eHealth Technology: Cross-Sectional Questionnaire Study
J Med Internet Res 2020;22(5):e14670
URL: http://www.jmir.org/2020/5/e14670/
doi:10.2196/14670
PMID:32348251

Original Paper

A User-Friendly, Web-Based Integrative Tool (ESurv) for Survival Analysis: Development and Validation Study

Kyoungjune Pak[1*], MD, PhD; Sae-Ock Oh[2*], MD, PhD; Tae Sik Goh[3*], MD, PhD; Hye Jin Heo[2], PhD; Myoung-Eun Han[2], PhD; Dae Cheon Jeong[4], MA; Chi-Seung Lee[5,6], PhD; Hokeun Sun[7], PhD; Junho Kang[8], MA; Suji Choi[8], BA; Soohwan Lee[8], BA; Eun Jung Kwon[8], BA; Ji Wan Kang[8], BA; Yun Hak Kim[2,8], MD, PhD

[1]Department of Nuclear Medicine, Pusan National University Hospital, Busan, Republic of Korea
[2]Department of Anatomy, School of Medicine, Pusan National University, Yangsan, Republic of Korea
[3]Department of Orthopaedic Surgery, Pusan National University Hospital, Busan, Republic of Korea
[4]Deloitte Analytics Group, Deloitte Consulting LLC, Seoul, Republic of Korea
[5]Biomedical Research Institute, Pusan National University Hospital, Busan, Republic of Korea
[6]Department of Biomedical Engineering, School of Medicine, Pusan National University, Busan, Republic of Korea
[7]Department of Statistics, Pusan National University, Busan, Republic of Korea
[8]Department of Biomedical Informatics, School of Medicine, Pusan National University, Yangsan, Republic of Korea
[*]these authors contributed equally

Corresponding Author:
Yun Hak Kim, MD, PhD
Department of Anatomy, School of Medicine
Pusan National University
49 Busandaehak-ro
Yangsan, 50612
Republic of Korea
Phone: 82 515108091
Email: yunhak10510@pusan.ac.kr

Abstract

Background: Prognostic genes or gene signatures have been widely used to predict patient survival and aid in making decisions pertaining to therapeutic actions. Although some web-based survival analysis tools have been developed, they have several limitations.

Objective: Taking these limitations into account, we developed ESurv (Easy, Effective, and Excellent Survival analysis tool), a web-based tool that can perform advanced survival analyses using user-derived data or data from The Cancer Genome Atlas (TCGA). Users can conduct univariate analyses and grouped variable selections using multiomics data from TCGA.

Methods: We used R to code survival analyses based on multiomics data from TCGA. To perform these analyses, we excluded patients and genes that had insufficient information. Clinical variables were classified as 0 and 1 when there were two categories (for example, chemotherapy: no or yes), and dummy variables were used where features had 3 or more outcomes (for example, with respect to laterality: right, left, or bilateral).

Results: Through univariate analyses, ESurv can identify the prognostic significance for single genes using the survival curve (median or optimal cutoff), area under the curve (AUC) with C statistics, and receiver operating characteristics (ROC). Users can obtain prognostic variable signatures based on multiomics data from clinical variables or grouped variable selections (lasso, elastic net regularization, and network-regularized high-dimensional Cox-regression) and select the same outputs as above. In addition, users can create custom gene signatures for specific cancers using various genes of interest. One of the most important functions of ESurv is that users can perform all survival analyses using their own data.

Conclusions: Using advanced statistical techniques suitable for high-dimensional data, including genetic data, and integrated survival analysis, ESurv overcomes the limitations of previous web-based tools and will help biomedical researchers easily perform complex survival analyses.

(J Med Internet Res 2020;22(5):e16084) doi:10.2196/16084

KEYWORDS

survival analysis; grouped variable selection; The Cancer Genome Atlas; web-based tool; user service

Introduction

The accumulation of large amounts of genomic data following the development of next-generation sequencing techniques is paving the way toward precision medicine [1-4]. In particular, gene expression profiles or signatures have been widely used to predict patient prognosis and assist in deciding therapeutic strategies for the treatment of various cancers [2,5-9].

Genomic data sets are highly variable, with that variability rising with an increasing number of patients, making it high-dimensional in nature. To efficiently link high-dimensional genomic and survival data, statisticians have developed grouped variable selection models, based on the Cox proportional hazards model, including the following: least absolute shrinkage and selection operator (lasso), elastic net regularization (elastic net), and network-regularized high-dimensional Cox-regression (Coxnet, hereon referred to as Net) [2,10-13]. Among these methods, Net has been found to have the fewest overfitting problems and the highest prediction performance in these applications, as it takes into consideration the complexities of biological networks [2,6,8,10,14].

Successfully identifying and verifying prognostic factors using big databases is essential in medical research, but this can be difficult for researchers who are unfamiliar with computer science. To address this unmet clinical need, some web-based survival analysis tools have been developed. Although these tools have some limitations, they have been used in some univariate analyses [15-18]. SurvExpress [19], PROGgene [20], and PrognoScan [21] are popular web-based survival analysis tools that calculate the statistical significance of a prognosis using only messenger RNA (mRNA) expression data [15-18].

The limitations of previous tools include the following: (1) The use of mRNA expression as a simple categorical value to provide Kaplan Meier curves for all patients, regardless of their characteristics. The use of a continuous variable like mRNA as a categorical factor can change the nature of the variable arbitrarily, resulting in serious errors. (2) These tools do not provide gene/variable signatures that are statistically better, in terms of predicting prognosis, than a single gene. (3) They do not consider cancer classifications like histological type. (4) Users cannot use their own data. (5) These tools do not provide high-quality images and tabular results. (6) Users cannot create a risk-scoring system by specifying the genes of interest.

To overcome these limitations, we developed ESurv (Easy, Effective, and Excellent Survival analysis tool [22]), which is an online web resource for identifying prognostic biomarkers in pan-cancer from The Cancer Genome Atlas (TCGA) or user data.

Methods

Processing Genetic and Clinical Information From Patients

We performed survival analyses based on multiomics and clinical data from TCGA (Table 1), obtained via Broad GDAC Firebrowse [23] and the GDC Data Portal [24-26]. We used level 3 RNAseq (RNA sequencing), miRNAseq (microRNA sequencing), and methylation array data. The criteria for exclusion were as follows: genes with 0 values of more than 10% (when the amount of missing data is greater than 10%, the results may be biased) [27], patients with insufficient overall survival information (survival time or status), and patients with paired normal tissue and metastasis samples in TCGA.

Table 1. Summary of the data available in the ESurv.

Cancers with available omics data	Messenger RNA (Yes/No)	MicroRNA (Yes/No)	Methylation (Yes/No)	Total patients, n
Acute myeloid leukemia	Yes	Yes	Yes	200
Adrenocortical carcinoma	Yes	Yes	Yes	92
Bladder urothelial carcinoma	Yes	Yes	Yes	412
Brain lower grade glioma	Yes	Yes	Yes	515
Breast invasive carcinoma	Yes	Yes	Yes	1097
Cervical and endocervical carcinoma	Yes	Yes	Yes	307
Cholangiocarcinoma	Yes	Yes	Yes	45
Colon adenocarcinoma	Yes	Yes	Yes	458
Esophageal carcinoma	Yes	Yes	Yes	185
Glioblastoma multiforme	Yes	No	Yes	595
Head and neck squamous cell carcinoma	Yes	Yes	Yes	528
Kidney chromophobe	Yes	Yes	Yes	113
Kidney renal clear cell carcinoma	Yes	Yes	Yes	537
Kidney renal papillary cell carcinoma	Yes	Yes	Yes	291
Liver hepatocellular carcinoma	Yes	Yes	Yes	377
Lung adenocarcinoma	Yes	Yes	Yes	522
Lung squamous cell carcinoma	Yes	Yes	Yes	504
Lymphoid neoplasm diffuse large B cell lymphoma	Yes	Yes	Yes	48
Mesothelioma	Yes	Yes	Yes	87
Ovarian serous cystadenocarcinoma	Yes	Yes	No	591
Pancreatic adenocarcinoma	Yes	Yes	Yes	185
Pheochromocytoma and paraganglioma	Yes	Yes	Yes	179
Prostate adenocarcinoma	Yes	Yes	Yes	499
Rectum adenocarcinoma	Yes	Yes	Yes	171
Sarcoma	Yes	Yes	Yes	261
Skin cutaneous melanoma	Yes	Yes	Yes	470
Stomach adenocarcinoma	Yes	Yes	Yes	443
Testicular germ cell tumor	Yes	Yes	Yes	134
Thymoma	Yes	Yes	Yes	124
Thyroid carcinoma	Yes	Yes	Yes	516
Uterine carcinosarcoma	Yes	Yes	Yes	57
Uterine corpus endometrial carcinoma	Yes	Yes	Yes	548
Uveal melanoma	Yes	Yes	No	80

Processing Clinical Variables in Net

Clinical variables (tumor stage, age, sex, cancer type, blast count, histologic grade, laterality, anatomic neoplasm subdivision, tumor tissue site, and human papillomavirus status) can be included in the Net depending on cancer variety, allowing for sophisticated analyses. Clinical variables were classified as 0 and 1 when there were two categories (for example, chemotherapy: no or yes), and dummy variables were applied when the clinical variable could fall into three or more categories (for example, laterality: right, left, or bilateral).

Grouped Variable Selections for Creating Gene/Variable Signatures

ESurv uses one of the following three methods: least absolute shrinkage and selection operator (lasso), elastic net regularization (elastic net), and network-regularized high-dimensional Cox-regression (Net) using the Coxnet package (version 0.2) in R [2,10,12,14]. In Net analysis, we

transformed the topological pathway information of large databases (KEGG [Kyoto Encyclopedia of Genes and Genomes], Biocarta, HumanCyc, Reactome, Panther, and NCI [National Cancer Institute]) into a gene network matrix using the graphite package (version 3.10) in R. Users can set the mixing parameter alpha, which decides the balance between lasso and ridge regression [10,14]. All grouped variable selections use 10-fold cross-validation. After variable selection, we calculated the prognostic score by multiplying the variable value by the regression coefficient.

Statistical Analysis

To determine the optimal cutoff value, preventing overoptimization, we used the maximal UNO's C-index and a 5-fold cross validation. For the Kaplan-Meier survival curves, patients were divided into two groups, high- and low-risk, based on specific gene expression parameters (median cutoff or optimal cutoff value), with a *P* value determined by a log-rank test. The C-index and area under the curve (AUC) value were used to evaluate the effect of specific variables on survival [28]. These results can be seen not only in all patients but also in patient subgroups based on sex or stage. The results were obtained using R packages survival (version 2.44-1.1), survMisc (version 0.5.5), coin (version 1.3-0), MASS (Modern Applied Statistics with S, version 7.3-51.4), edgeR (version 3.24.3), and survAUC (version 1.0-5). All graphical outputs from ESurv were plotted using the plotly (version 4.9.0) R package. All data cleaning and statistical analyses in this study were performed using R statistical software (version 3.6.0, R Core Team, R Foundation for Statistical Computing).

Implementation

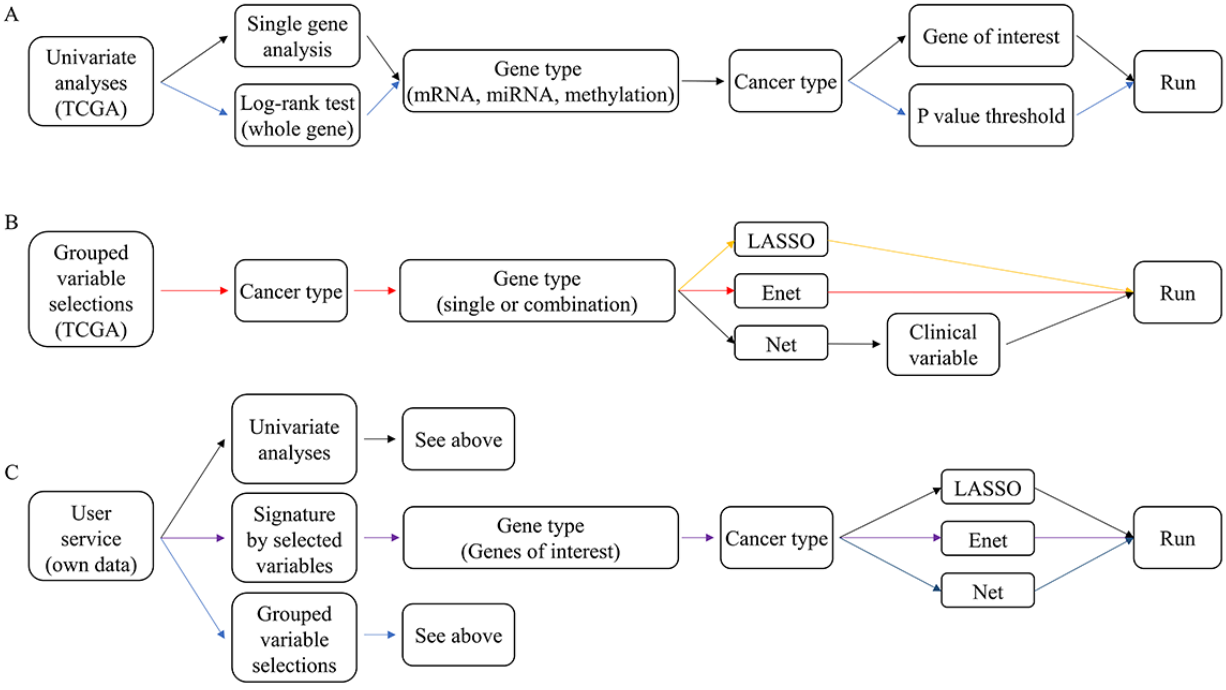
The ESurv web server implements AngularJS with HTML5 to display analyzed data through a web query interface. The results of these analyses were calculated on demand on a backend server running Java Servlet in conjunction with the R statistical program.

Results

Running ESurv

Details of the running procedure for ESurv are described in Figure 1. First, users can choose one of three methods: univariate survival analyses, grouped variable selections, or user service.

Figure 1. The running procedure of ESurv.



After choosing univariate survival analysis, users can select single gene analysis or log-rank test of whole genomes. In single gene analysis, users select the type of cancer, gene data type (mRNA, microRNA [miRNA], or methylation), gene of interest, and time for the receiver operating characteristics (ROC) curve, in that order. For log-rank testing of whole genomes, users choose the type of cancer, gene data type (mRNA, miRNA, or methylation), time for the ROC curve, and *P* value threshold, in that order.

If users instead choose grouped variable selections, they must select the type of cancer, gene data (mRNA, miRNA, methylation, or integrative analysis), grouped variable selection method (lasso, elastic net, or Net), time for the ROC curve, and alpha, in that order. Alpha decides the balance between ridge and lasso penalties; the larger the alpha, the closer to lasso (alpha=1), and the fewer variables are chosen. If the users select Net, they can include clinical variables in grouped variable selection.

In the user service, users can perform univariate or grouped variable selections after uploading their own data. Instructions on uploading data are detailed in the manual. Once data are uploaded, all the abovementioned analyses can be performed.

Finally, when developing gene signatures using selected variables, users should choose the cancer type, genes of interest, and time for the ROC curve, in that order.

Univariate Analysis: Single Gene Analysis in Pan-Cancer

ESurv determines the prognostic significance of single genes as categorical (Kaplan-Meier curve with median or optimal cutoff values) or continuous variables (C-index and AUC values at specific time points) in various subgroups including sex and cancer stage (early and late as defined by the American Joint Committee on Cancer). To complement the results from the categorical variables, the C-index and AUC values were calculated as continuous variables. For example, we performed survival analysis using mRNA (*SLC2A1*) as a biomarker in bladder urothelial carcinoma (n=403; Figure 2). As shown in Figure 2A and B, the discrimination power with optimal cutoff values is much better than that of analyses completed using median cutoff values in all patients. Users easily obtain survival curves in subgroups as well as for all patients. In the time-dependent ROC curve analyses, users can identify the AUC value based on the follow-up time (Figure 2C). ROC curves at selected times can all be calculated (Figure 2D).

Figure 2. An example of mRNA-based survival analysis. Expression levels of genes are classified as low or high (blue or red lines, respectively) based on the comparison of their optimal (A) and median cut-off values (B). (C) Time-dependent area under the curve (AUC) for each of these subgroups. (D) Receiver operating characteristics (ROC) curves for selected years in each of these subgroups.

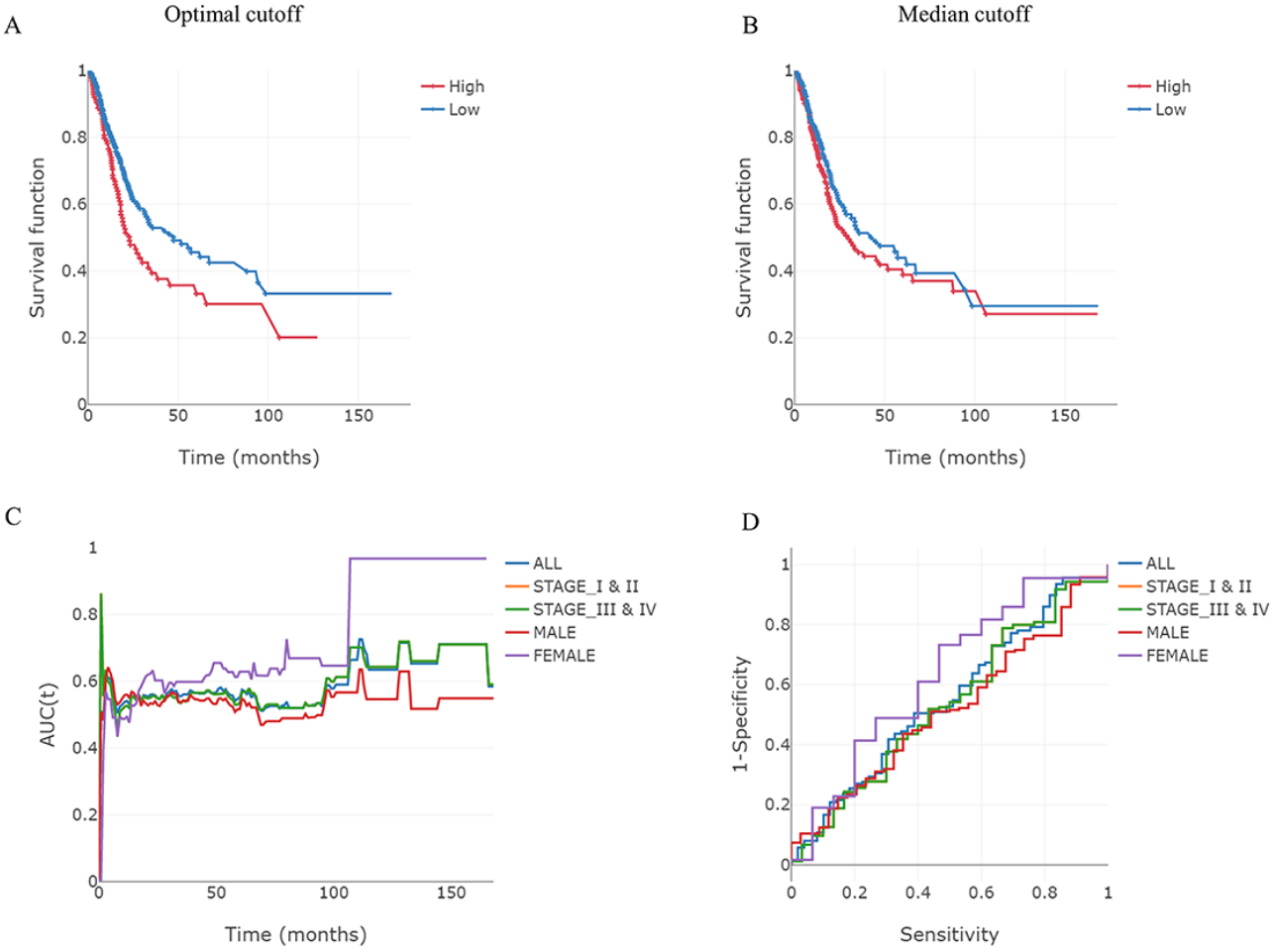
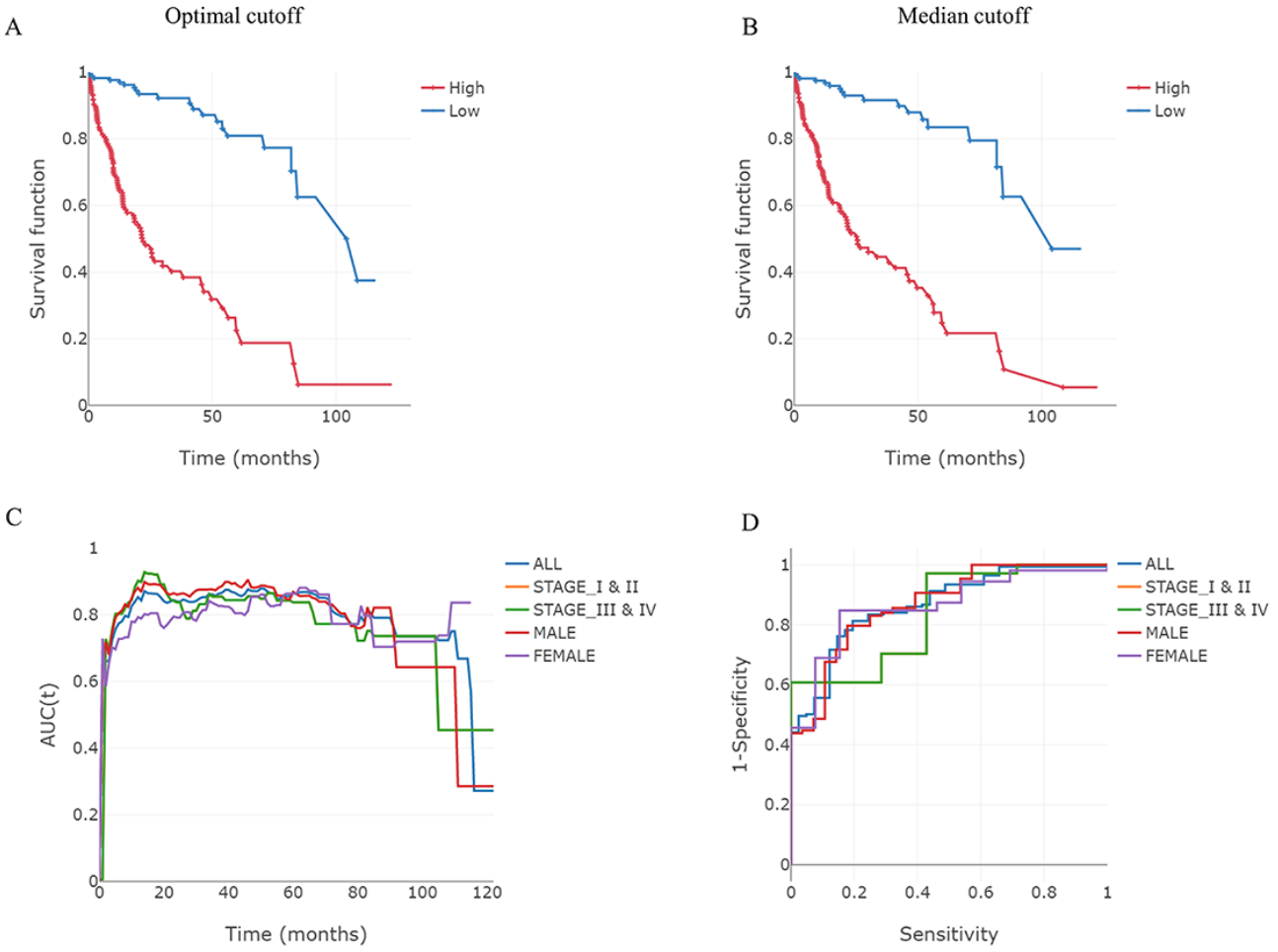


Figure 3. An example of survival analysis using a variable signature. Expression levels of genes were classified as either low or high (blue or red lines, respectively) based on a comparison of their optimal (A) and median cut-off values (B). (C) Time-dependent area under the curve (AUC) for each of the subgroups. (D) Receiver operating characteristics (ROC) curves for selected years in each of these subgroups.



Univariate Analysis: Log-Rank Test of Whole Genomes

If the users want to calculate the prognostic value of all gene variants in a particular cancer, they can choose to do a log-rank test for the whole genome. Here, we performed a log-rank test of whole genomes in colon adenocarcinoma with a $P<.05$. ESurv provided the results (gene name, C-index, AUC value at user-selected time, P value of log-rank test) as an Excel file (Multimedia Appendix 1).

Grouped Variable Selections

Grouped variable selection methods have been developed to take advantage of advances in biological technology, and can be used in statistical models to accurately predict the prognosis for a patient [2,10]. To select a prognostic gene set from high-dimensional data, it is necessary to select and minimize the number of variables in a systematic and statistically sound manner. ESurv provides representative grouped variable selection methods (lasso, elastic net, and Net), which can reduce the number of variables by considering the relationships between them. To develop a risk-scoring system, we used a linear combination of the expression values and regression coefficients of the selected variables. Users can develop a powerful prognostic signature in pan-cancer using any combination of these options. To illustrate this, we performed Net using all the clinical variables in hepatocellular carcinoma (n=307; Figure 3). Selected variables and regression coefficients can be downloaded from the summary tab in ESurv as Excel files. The results for all possible subgroups were recorded, just as in the case of single gene analysis. Outputs for the analysis can be in the form of Kaplan-Meier curves, time-dependent ROC curves, and AUC values at specific time points (Figure 3).

User Service

Users can conduct univariate analysis, log-rank tests, and grouped variable selections by uploading their own data (Figure 1C). The manual is provided in Multimedia Appendices 2 and 3. Users can generate all survival results, as described above, using their own data sets. To protect data integrity, user-derived data sets are password-protected and connected to a unique user ID. In addition, as users refine their needs, they can request additional survival analysis packages via email. The most-requested package will be added to ESurv the following year.

Discussion

Principal Findings

To overcome the limitations of existing survival analysis tools, we developed a web-based user-friendly tool called ESurv. Kaplan-Meier curves are a common method of conducting survival analysis in the medical field, which involves

categorizing patients into groups based on their risk profile. There is no clear criterion to classify continuous variables, like gene expression, in these categorical analysis methods. For this reason, median and quartile cutoff values used in previous tools may miss the prognostic significance of individual genes. As shown in Figure 2B, the median cutoff is not a suitable parameter for risk stratification, whereas optimal cutoff is a good parameter for risk stratification. Additionally, survival analysis as a continuous variable should be accompanied by these results to evaluate the prognostic significance of the gene of interest. Researchers can identify the prognostic significance of a gene (mRNA, miRNA, and methylation) as both a categorical and continuous variable in ESurv.

Prognosis using multiple genes yields superior results compared to using a single gene. There are a number of ways to select gene signatures, but among these, genes selected using grouped variable selection have proven to be the most versatile and reproducible [2,5,9,10,13,29,30]. Grouped variable selection methods select and shrink variables from high-dimensional data sets while considering multicollinearity, which is especially valuable when considering biological pathways, making Net options ideal for these applications. Despite this, grouped variable selection has not been applied to many studies because of the difficulty surrounding its computer programming (R, Python, and Matlab). To address this gap, we added grouped variable selection methods (lasso, elastic net, and Net) to ESurv. When selecting a prognostic gene set from high-dimensional data, researchers have to minimize the number of variables. In order to reduce variables effectively, we must provide as much information as possible linking genes and their relevant pathways [2,10]. It is easier to complete external validation with Net than with other methods because it performs variable selection using information about each gene, derived from databases hosting information pertaining to genetic pathways (Reactome, HumanCyc, KEGG, Biocarta, NCI, and Panther) [2,10,29,30]. Like with the univariate analyses, users can obtain the results as continuous and categorical data. ESurv is the first web-based tool to provide grouped variable selection using multiomics data.

The prognosis of patients may vary based on clinical information, such as sex and stage [9,31-35]. For these reasons, subgroup analysis is required to identify stage- and sex-specific prognostic genes. ESurv shows the results of survival analyses by taking into consideration tumor heterogeneity based on several classifications (cancer type, stages, and sex).

Limitations and Future Work

This software does have some limitations that will be addressed upon further development. Here we used only one cancer database but there are many more, which we plan to add to ESurv as we continue to develop the software. Although users can upload their own data, this still requires users to be computer savvy; this will be addressed in future versions of the software. In addition, ESurv currently only accesses cancer databases, but this type of analysis is valuable in other diseases, including several vascular and degenerative diseases; we aim to add these as well. Finally, ESurv is not exhaustive in its analyses, but it is possible for users to request additional survival analysis packages for R via email. We will then select the most requested package and add it to ESurv on an annual basis.

Conclusions

The most important functionality provided by ESurv is that users can use this software to analyze their own data. As more medical data is produced, the demand for survival analyses increases. Analysis of data created by individual institutions is as important as big data analysis, but there have been no survival analysis tools available to conduct this type of analysis. Using advanced statistical methods and comprehensive survival analyses, ESurv overcomes the limitations of previous tools, and allows users to work on their own data sets. We strongly believe that ESurv is an ideal tool to meet the growing demand for increased survival analysis in both small and large data sets.

Acknowledgments

This work was supported by grants from the Medical Research Center (MRC) Program and the Basic Science Research Program through the National Research Foundation of Korea (NRF) grant funded by the government of Korea (NRF-2018R1A5A2023879 and NRF-2019R1A2B5B01070163). This study was supported by a Biomedical Research Institute Grant (2018B032) from the Pusan National University Hospital.

Authors' Contributions

KP and TSG designed and wrote the manuscript. HJH, MEH, CSL, JK, SJC, and SL collected and preprocessed the data. DCJ and HS corrected the error in the algorithm. SOO and YHK designed and supervised the project. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The univariate results of whole significant genes.
[XLSX File (Microsoft Excel File), 339 KB - jmir_v22i5e16084_app1.xlsx]

Multimedia Appendix 2
The manual of univariate analyses.
[DOCX File , 67 KB - jmir_v22i5e16084_app2.docx]

Multimedia Appendix 3
The manual of grouped variable selection analyses.
[DOCX File , 72 KB - jmir_v22i5e16084_app3.docx]

References

1. Friedman AA, Letai A, Fisher DE, Flaherty KT. Precision medicine for cancer with next-generation functional diagnostics. Nat Rev Cancer 2015 Dec;15(12):747-756 [FREE Full text] [doi: 10.1038/nrc4015] [Medline: 26536825]
2. Kim YH, Jeong DC, Pak K, Goh TS, Lee C, Han M, et al. Gene network inherent in genomic big data improves the accuracy of prognostic prediction for cancer patients. Oncotarget 2017 Sep 29;8(44):77515-77526 [FREE Full text] [doi: 10.18632/oncotarget.20548] [Medline: 29100405]
3. Han M, Kim J, Kim GH, Park SY, Kim YH, Oh S. SAC3D1: a novel prognostic marker in hepatocellular carcinoma. Sci Rep 2018 Oct 23;8(1):15608 [FREE Full text] [doi: 10.1038/s41598-018-34129-9] [Medline: 30353105]
4. Shaikh AR, Butte AJ, Schully SD, Dalton WS, Khoury MJ, Hesse BW. Collaborative biomedicine in the age of big data: the case of cancer. J Med Internet Res 2014 Apr 07;16(4):e101 [FREE Full text] [doi: 10.2196/jmir.2496] [Medline: 24711045]
5. Chen H, Yu S, Chen C, Chang G, Chen C, Yuan A, et al. A five-gene signature and clinical outcome in non-small-cell lung cancer. N Engl J Med 2007 Jan 04;356(1):11-20. [doi: 10.1056/NEJMoa060096] [Medline: 17202451]
6. Zemmour C, Bertucci F, Finetti P, Chetrit B, Birnbaum D, Filleron T, et al. Prediction of early breast cancer metastasis from DNA microarray data using high-dimensional cox regression models. Cancer Inform 2015;14(Suppl 2):129-138 [FREE Full text] [doi: 10.4137/CIN.S17284] [Medline: 25983547]
7. Nault J, De Reyniès A, Villanueva A, Calderaro J, Rebouissou S, Couchy G, et al. A hepatocellular carcinoma 5-gene score associated with survival of patients after liver resection. Gastroenterology 2013 Jul;145(1):176-187. [doi: 10.1053/j.gastro.2013.03.051] [Medline: 23567350]
8. Zhang J, Song W, Chen Z, Wei J, Liao Y, Lei J, et al. Prognostic and predictive value of a microRNA signature in stage II colon cancer: a microRNA expression analysis. Lancet Oncol 2013 Dec;14(13):1295-1306. [doi: 10.1016/S1470-2045(13)70491-1] [Medline: 24239208]
9. Lau SK, Boutros PC, Pintilie M, Blackhall FH, Zhu C, Strumpf D, et al. Three-gene prognostic classifier for early-stage non small-cell lung cancer. J Clin Oncol 2007 Dec 10;25(35):5562-5569. [doi: 10.1200/JCO.2007.12.0352] [Medline: 18065728]
10. Sun H, Lin W, Feng R, Li H. Network-Regularized High-Dimensional Cox Regression for Analysis of Genomic Data. Stat Sin 2014 Jul;24(3):1433-1459 [FREE Full text] [doi: 10.5705/ss.2012.317] [Medline: 26316678]
11. Musoro JZ, Zwinderman AH, Puhan MA, ter Riet G, Geskus RB. Validation of prediction models based on lasso regression with multiply imputed data. BMC Med Res Methodol 2014 Oct 16;14:116 [FREE Full text] [doi: 10.1186/1471-2288-14-116] [Medline: 25323009]
12. Tibshirani R. The lasso method for variable selection in the Cox model. Stat Med 1997 Feb 28;16(4):385-395. [Medline: 9044528]
13. Fulton L, Kruse CS. Hospital-Based Back Surgery: Geospatial-Temporal, Explanatory, and Predictive Models. J Med Internet Res 2019 Oct 29;21(10):e14609 [FREE Full text] [doi: 10.2196/14609] [Medline: 31663856]
14. Wu Y. ELASTIC NET FOR COX'S PROPORTIONAL HAZARDS MODEL WITH A SOLUTION PATH ALGORITHM. Stat Sin 2012;22:27-294 [FREE Full text] [doi: 10.5705/ss.2010.107] [Medline: 23226932]
15. Aguirre-Gamboa R, Gomez-Rueda H, Martínez-Ledesma E, Martínez-Torteya A, Chacolla-Huaringa R, Rodriguez-Barrientos A, et al. SurvExpress: an online biomarker validation tool and database for cancer gene expression data using survival analysis. PLoS One 2013;8(9):e74250 [FREE Full text] [doi: 10.1371/journal.pone.0074250] [Medline: 24066126]
16. Goswami CP, Nakshatri H. PROGgene: gene expression based survival analysis web application for multiple cancers. J Clin Bioinforma 2013 Oct 28;3(1):22 [FREE Full text] [doi: 10.1186/2043-9113-3-22] [Medline: 24165311]
17. Goswami CP, Nakshatri H. PROGgeneV2: enhancements on the existing database. BMC Cancer 2014 Dec 17;14:970 [FREE Full text] [doi: 10.1186/1471-2407-14-970] [Medline: 25518851]
18. Mizuno H, Kitada K, Nakai K, Sarai A. PrognoScan: a new database for meta-analysis of the prognostic value of genes. BMC Med Genomics 2009 Apr 24;2:18 [FREE Full text] [doi: 10.1186/1755-8794-2-18] [Medline: 19393097]
19. SurvExpress. URL: http://bioinformatica.mty.itesm.mx:8080/Biomatec/SurvivaX.jsp [accessed 2020-04-08]
20. PROGgene. URL: http://www.compbio.iupui.edu/proggene [accessed 2020-04-08]
21. PrognoScan. URL: http://dna00.bio.kyutech.ac.jp/PrognoScan/index.html [accessed 2020-04-08]
22. ESurv. URL: https://easysurv.net [accessed 2020-04-08]
23. Broad GDAC Firebrowse. URL: https://firebrowse.org [accessed 2020-04-08]

24. GDC Data Portal. URL: https://portal.gdc.cancer.gov/ [accessed 2020-04-08]
25. Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet 2013 Oct;45(10):1113-1120 [FREE Full text] [doi: 10.1038/ng.2764] [Medline: 24071849]
26. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, Cancer Genome Atlas Research Network, et al. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. Cell 2018 Apr 05;173(2):400-416.e11 [FREE Full text] [doi: 10.1016/j.cell.2018.02.052] [Medline: 29625055]
27. Bennett D. How can I deal with missing data in my study? Aust N Z J Public Health 2001 Oct;25(5):464-469. [Medline: 11688629]
28. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. Stat Med 2011 May 10;30(10):1105-1117 [FREE Full text] [doi: 10.1002/sim.4154] [Medline: 21484848]
29. Goh TS, Lee JS, Il Kim J, Park YG, Pak K, Jeong DC, et al. Prognostic scoring system for osteosarcoma using network-regularized high-dimensional Cox-regression analysis and potential therapeutic targets. J Cell Physiol 2019 Aug;234(8):13851-13857. [doi: 10.1002/jcp.28065] [Medline: 30604867]
30. Pak K, Kim YH, Suh S, Goh TS, Jeong DC, Kim SJ, et al. Development of a risk scoring system for patients with papillary thyroid cancer. J Cell Mol Med 2019 Apr;23(4):3010-3015 [FREE Full text] [doi: 10.1111/jcmm.14208] [Medline: 30729678]
31. Benedix F, Kube R, Meyer F, Schmidt U, Gastinger I, Lippert H, Colon/Rectum Carcinomas (Primary Tumor) Study Group. Comparison of 17,641 patients with right- and left-sided colon cancer: differences in epidemiology, perioperative course, histology, and survival. Dis Colon Rectum 2010 Jan;53(1):57-64. [doi: 10.1007/DCR.0b013e3181c703a4] [Medline: 20010352]
32. Hansen IO, Jess P. Possible better long-term survival in left versus right-sided colon cancer - a systematic review. Dan Med J 2012 Jun;59(6):A4444. [Medline: 22677242]
33. Kim S, Paik HY, Yoon H, Lee JE, Kim N, Sung M. Sex- and gender-specific disparities in colorectal cancer risk. World J Gastroenterol 2015 May 07;21(17):5167-5175 [FREE Full text] [doi: 10.3748/wjg.v21.i17.5167] [Medline: 25954090]
34. Lieberman DA, Williams JL, Holub JL, Morris CD, Logan JR, Eisen GM, et al. Race, ethnicity, and sex affect risk for polyps >9 mm in average-risk individuals. Gastroenterology 2014 Aug;147(2):351-8; quiz e14–5 [FREE Full text] [doi: 10.1053/j.gastro.2014.04.037] [Medline: 24786894]
35. Tian S. Identification of Subtype-Specific Prognostic Genes for Early-Stage Lung Adenocarcinoma and Squamous Cell Carcinoma Patients Using an Embedded Feature Selection Algorithm. PLoS One 2015;10(7):e0134630 [FREE Full text] [doi: 10.1371/journal.pone.0134630] [Medline: 26226392]

Abbreviations

AUC: area under the curve
Elastic net: elastic net regularization
ESurv: Easy, Effective, and Excellent Survival analysis tool
Lasso: least absolute shrinkage and selection operator
miRNA: microRNA
mRNA: messenger RNA
Net: network-regularized high-dimensional Cox-regression
ROC: receiver operating characteristics
TCGA: The Cancer Genome Atlas

Edited by G Eysenbach; submitted 02.09.19; peer-reviewed by S Suh, N Borcherding; comments to author 22.10.19; revised version received 13.11.19; accepted 25.03.20; published 05.05.20.

Please cite as:
Pak K, Oh SO, Goh TS, Heo HJ, Han ME, Jeong DC, Lee CS, Sun H, Kang J, Choi S, Lee S, Kwon EJ, Kang JW, Kim YH
A User-Friendly, Web-Based Integrative Tool (ESurv) for Survival Analysis: Development and Validation Study
J Med Internet Res 2020;22(5):e16084
URL: https://www.jmir.org/2020/5/e16084
doi:10.2196/16084
PMID:32369034

Original Paper

Using Machine Learning to Predict Early Onset Acute Organ Failure in Critically Ill Intensive Care Unit Patients With Sickle Cell Disease: Retrospective Study

Akram Mohammed[1], PhD; Pradeep S B Podila[2], PhD, MHA, MS; Robert L Davis[1], MD, MPH; Kenneth I Ataga[3], MD; Jane S Hankins[4], MD, MS; Rishikesan Kamaleswaran[5], PhD

[1]Center for Biomedical Informatics, University of Tennessee Health Science Center, Memphis, TN, United States

[2]Faith and Health Division, Methodist Le Bonheur Healthcare, Memphis, TN, United States

[3]Center for Sickle Cell Disease, University of Tennessee Health Science Center, Memphis, TN, United States

[4]Department of Hematology, St Jude Children's Research Hospital, Memphis, TN, United States

[5]Department of Biomedical Informatics, Emory University School of Medicine, Atlanta, GA, United States

Corresponding Author:
Rishikesan Kamaleswaran, PhD
Department of Biomedical Informatics
Emory University School of Medicine
WMB, 101 Woodruff Circle, Suite 4127
Atlanta, GA, 30322
United States
Phone: 1 (901) 462 6908
Email: rkamaleswaran@emory.edu

Abstract

Background: Sickle cell disease (SCD) is a genetic disorder of the red blood cells, resulting in multiple acute and chronic complications, including pain episodes, stroke, and kidney disease. Patients with SCD develop chronic organ dysfunction, which may progress to organ failure during disease exacerbations. Early detection of acute physiological deterioration leading to organ failure is not always attainable. Machine learning techniques that allow for prediction of organ failure may enable early identification and treatment and potentially reduce mortality.

Objective: The aim of this study was to test the hypothesis that machine learning physiomarkers can predict the development of organ dysfunction in a sample of adult patients with SCD admitted to intensive care units (ICUs).

Methods: We applied diverse machine learning methods, statistical methods, and data visualization techniques to develop classification models to distinguish SCD from controls.

Results: We studied 63 sequential SCD patients admitted to ICUs with 163 patient encounters (mean age 30.7 years, SD 9.8 years). A subset of these patient encounters, 22.7% (37/163), met the sequential organ failure assessment criteria. The other 126 SCD patient encounters served as controls. A set of signal processing features (such as fast Fourier transform, energy, and continuous wavelet transform) derived from heart rate, blood pressure, and respiratory rate was identified to distinguish patients with SCD who developed acute physiological deterioration leading to organ failure from patients with SCD who did not meet the criteria. A multilayer perceptron model accurately predicted organ failure up to 6 hours before onset, with an average sensitivity and specificity of 96% and 98%, respectively.

Conclusions: This retrospective study demonstrated the viability of using machine learning to predict acute organ failure among hospitalized adults with SCD. The discovery of salient physiomarkers through machine learning techniques has the potential to further accelerate the development and implementation of innovative care delivery protocols and strategies for medically vulnerable patients.

(J Med Internet Res 2020;22(5):e14693) doi:10.2196/14693

KEYWORDS

multiple organ failure; sickle cell disease; machine learning; electronic medical record; hematology

Introduction

Background

Sickle cell disease (SCD), one of the most common genetic disorders, affects millions across the globe [1]. It was the first monogenic disorder to be characterized at the molecular level. It is characterized by the presence of abnormal hemoglobin S, which, under hypoxic conditions, causes sickling of red blood cells, resulting in tissue and organ damage. Among an array of complications afflicting patients with SCD, the most devastating is major organ failure, including pulmonary failure, end-stage renal disease, stroke, and heart failure [1]. A 4-decade observational study reported that, by the fifth decade of life, up to half of all patients with SCD had documented irreversible organ damage [2]. Organ dysfunction may manifest or worsen during hospitalizations, when disease complications arise. Thus, therapy supplemented by predictive analytics can potentially improve the outcomes of patients with SCD [3]. Early diagnosis of acute organ dysfunction may allow for early intervention, thereby preventing or reducing the severity of organ failure, particularly during hospitalization for acute complications.

Early recognition of organ failure may [4,5] thereby enable clinicians to provide targeted therapies to improve outcomes. Various scoring methods have been developed for qualifying organ dysfunction, including Acute Physiology and Chronic Health Evaluation [6], Multi-Organ Dysfunction Score [7], *quick* Sequential Organ Failure Assessment [8] and Sequential Organ Failure Assessment (SOFA) [9]. The SOFA is a mortality prediction score that is based on the degree of dysfunction of six organ systems. The score is calculated at admission and every 24 hours until discharge, using the worst parameters measured during the previous 24 hours. Compared with other scoring methods, the use of SOFA allowed us to retrospectively quantify both the number and severity of individual organ dysfunction.

Objectives

In this retrospective study, we used serial calculations of SOFA to identify the onset of organ failure and then used physiomarkers in machine learning models to predict organ failure for patients with SCD presenting with a severe, acute painful crisis. Our hypothesis was that physiomarkers [10] identified by machine learning methods can be used to predict organ failure.

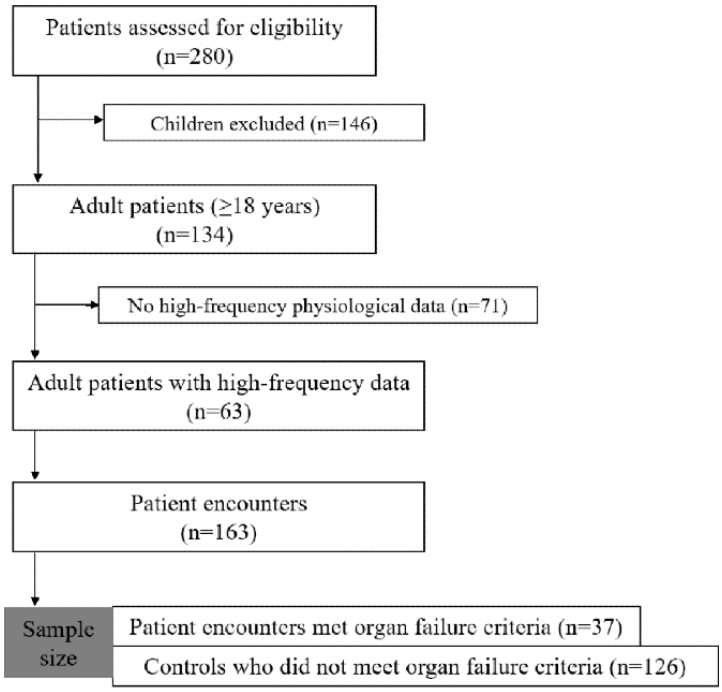
Methods

Cohort

Continuous physiologic data were collected on 134 adult subjects with SCD admitted to intensive care units (ICUs) at Methodist Le Bonheur Healthcare hospitals, Memphis, Tennessee, United States, between June 2017 and March 2018. Patients were retrospectively identified using a discharge International Classification of Diseases, Tenth Revision (ICD-10) code of D57.*. Of the 134 unique patients, 71 patients who did not have at least 24 hours of continuous physiologic data were excluded from the analysis. We studied patients who had at least 24 hours of continuous high-frequency physiologic data available before the time of organ failure onset (identified using SOFA criteria). A total of 63 unique adult subjects who had SCD (discharge ICD-10 code of D57.*) and were admitted to the medical, surgical, neurological, and cardiac ICUs and had continuous physiologic data available were retrospectively identified and included in the study. These 63 patients had 163 encounters (Figure 1). Of the 163 encounters, 37 patient encounters corresponding to 29 unique patients met organ failure criteria. The inclusion and exclusion criteria are summarized in Figure 1. Demographic and clinical data on cases and controls were collected from the electronic medical record (EMR) using Cerner's Web Intelligence reporting module (Cerner Health Facts). Each patient admission was considered as a separate patient encounter if the interval between admissions was at least for 1 month. The principal or admit diagnosis was identified using the ICD-10 codes. Patients with organ failure at admission were excluded from the analysis.

High-frequency physiologic data were collected at the frequency of once per minute from the time of admission until discharge. A total of 5 physiologic characteristics were used in the analysis, including heart rate (HR), respiratory rate (RR), systolic blood pressure (SBP), diastolic blood pressure (DBP), and mean blood pressure (MBP). These vital signs were selected because they were always obtained in all patients admitted to ICUs, and these can be used in identifying organ failure using a minimal set of physiologic data. The main outcome of organ failure was the failure of at least one organ or system (cardiovascular, liver, respiratory, coagulation, central nervous system, or renal), and cases were defined as patients meeting an increase in a serially calculated SOFA criteria by at least one score within a 24-hour rolling window, from admission till discharge. Event time (t_{onset}) was recorded as the earliest time stamp of every occurrence of organ failure, as defined independently by SOFA for each patient. Patients without 24 hours of physiological data before the t_{onset} (for cases and controls) were excluded. To normalize our prediction time horizon, we created relative alignments of time windows, pivoted to t_{onset}. This study was approved by The University of Tennessee Health Science Center and Methodist Le Bonheur Hospital Institutional Review Boards, and it was performed in compliance with the ethical principles for medical research involving human subjects from the Declaration of Helsinki.

Figure 1. Consolidated Standards for Reporting of Trials diagram describing the study cohort.



Feature Extraction and Feature Selection

We utilized Python libraries for extracting features from each of the physiological data streams, including a combination of temporal, frequency, and statistical features (Multimedia Appendix 1) [11]. Features were derived from six overlapping 3-hour time intervals, with a stride of 1 hour, from 1 to 4 hours to 6 to 9 hours before organ failure, so that we could build predictive models at different times before organ failure. For controls (patients with SCD not developing organ failure), we extracted these same features from 3-hour intervals before a random time period (identified by time stamp) during their ICU stay. The detailed feature extraction for organ failure cases and controls is shown in Multimedia Appendix 2. We performed feature selection using multiple null hypothesis testing, using the Benjamini-Yekutieli procedure and Mann-Whitney *U* test. Finally, we ranked the features using random forest (RF) [12] feature importance algorithm and applied various feature thresholds to select subsets of features that were most discriminatory among cases and controls.

Machine Learning Algorithms

Multilayer perceptron (MLP), support vector machine (SVM), RF, and logistic regression (LR) methods were used for building classification models. These methods were adopted because of their successful applications to medical datasets for disease classification [13-18].

An MLP is a deep, feed-forward neural network comprising an input layer, an output layer, and at least two or more hidden layers [19]. MLP has been used in a variety of applications, including electroencephalogram signal classification [15], heart disease diagnosis [20], ovarian tumor classification [21], and continuous speech recognition [22]. The MLP architecture used in this study comprised five hidden layers of 512, 256, 128, 64, and 16 neurons. We applied batch normalization [23] before activation, using the rectified linear unit. To avoid overfitting, we further imposed a dropout [24] ratio of 0.3. The output layer performed binary classification using the sigmoid activation, and our loss function used the Adam optimizer.

The SVM is a multivariate machine learning approach for classifying samples through a pattern recognition analysis [25]. SVM aims to find the best hyperplane that separates all data points of one class compared with those of another class. We used a radial basis function as the kernel parameter for model building.

The RF classifier is well suited to the classification of medical data because of the following advantages: (1) it performs embedded feature selection, (2) it incorporates interactions between predictors, (3) it allows the algorithm to learn both simple and complex classification functions accurately, and (4) it is applicable to both binary and multicategory classification tasks [12]. On the basis of the out-of-bag error [26], we identified 500 trees in the RF models as the optimal number of trees.

LR can be used as a machine learning method used to predict the value of a binary variable based on its relationship with predictor variables [27]. The *P* value for statistical testing of variable significance for *inclusion-in* and *exclusion-from* the model was set to $P=.05$ and $P=.10$, respectively, and LIBLINEAR solver was used for the optimization function.

Statistical Analysis and Machine Learning Framework

Python scikit-learn machine learning library [28] was used for calculating descriptive statistical measures, for feature selection, and for building machine learning classifiers. Bivariate LR, bootstrap, and Bayesian bootstrap (adjusted for weights) were used to assess the predictability of the features generated for predicting organ failure [29]. We used nonparametric Kruskal-Wallis statistical tests to analyze the difference among the five physiological signals (DBP, SBP, MBP, HR, and RR).

Each of these five signals had six feature measurements, namely, mean, energy ratio, fast Fourier transform, linear trend, quantile, and continuous wavelet transform.

Cross Validation

Models were developed from the distinct time intervals and tested on patients who were not included in the training of the model. For model selection and accuracy estimation, we used 5-fold cross validation [30]. This technique divides data into five equal and discrete folds and uses four folds for model generation, whereas predictions are generated and evaluated using the remaining single fold. This step is subsequently repeated five times, so each fold is tested against the other four folds. We further ran each of these 5-fold cross-validation models 10 times by shuffling the data in each iteration and averaged the performance metrics from all iterations to reduce bias.

Results

Patient Characteristics

Tables 1 and 2 outline the descriptive-level characteristics of demographics and clinical characteristics of 163 encounters. Four patients with organ failure at admission were omitted from further analysis and their data are not shown in Table 1.

Table 1. Encounter-level demographics and principal diagnosis of patients in the overall cohort (n=163).

Variable[a]	Total cohort	Organ failure (yes)	Organ failure (no)	*P* value
Total sample, n (%)	163 (100.0)	37 (22.7)	126 (77.3)	N/A[b]
Age (years), mean (SD)	30.7 (9.8)	35.2 (12.9)	29.3 (8.3)	.01[c]
Female, n (%)	87 (53.4)	24 (64.9)	63 (50.0)	.11
African American, n (%)	163 (100.0)	37 (100.0)	126 (100.0)	N/A
Admit diagnosis, n (%)[d]				
Vaso-occlusive event (pain or acute chest syndrome)	130 (79.8)	23 (62.2)	107 (84.9)	.003[c]
Nonvaso-occlusive crises pain	7 (4.3)	4 (10.8)	3 (2.4)	.05[c]
Infection/sepsis	7 (4.3)	2 (5.4)	5 (4.0)	.66
Other[e]	19 (11.6)	8 (21.6)	11 (8.7)	.03[c]

[a]For continuous variables, independent *t* test was used; for categorical variables, Chi-square test of independence was used. Fisher exact test was used for variables with cell counts of less than 5.

[b]N/A: Not applicable.

[c]Statistically significant at *P*=.05.

[d]The admit diagnoses are based on the International Classification of Diseases, Tenth Revision, Clinical Modification codes at the admission time.

[e]Other category includes respiratory distress, sickle cell disease without crisis, diabetes complications (diabetic ketoacidosis/hyperglycemia), pneumonia, myocardial infarction, hematemesis, cough, and deep venous thrombosis.

The mean age of the patient encounters in the cohort was 30.7 years (9.8 years); all patients were African American, and there were more females, 53.4% (87/163), than males, 46.6% (76/163). Admit diagnoses of vaso-occlusive event (pain or acute chest syndrome; 130/163, 79.8%), nonvaso-occlusive crises pain (7/163, 4.3%), and infection/sepsis (7/163, 4.3%) were common (Table 1). Both vaso-occlusive and nonvaso-occlusive events significantly altered between the patient with *organ failure* and *no organ failure* groups. Patients with organ failure had longer hospital stays (3.2 additional hospital days; *P*=.03) than controls, and they had higher severity of illness (*P*<.001) and risk of mortality (*P*<.001; Table 2).

Table 2. Encounter-level clinical characteristics of patients in the overall cohort (n=163).

Variable[a]	Total cohort	Organ failure (yes)	Organ failure (no)	*P* value
Encounters, n (%)	163 (100.0)	37 (22.7)	126 (77.3)	—[b]
Encounter through emergency department, n (%)	134 (82.2)	33 (89.2)	101 (80.2)	.33
Length of stay (days), mean (SD)	5.3 (4.7)	7.7 (8.4)	4.5 (2.5)	.03[c]
APR-DRG[d] severity of illness, n (%)				**<.001**
Minor	52 (37.1)	7 (19.4)	45 (43.2)	
Moderate	40 (28.6)	7 (19.4)	33 (31.7)	
Major	40 (28.6)	15 (41.7)	25 (24.0)	
Extreme	8 (5.7)	7 (19.4)	1 (1.0)	
APR-DRG risk of mortality, n (%)				**<.001**
Minor	93 (66.4)	14 (38.9)	79 (76.0)	
Moderate	26 (18.6)	6 (16.7)	20 (19.2)	
Major	12 (8.6)	8 (22.2)	4 (3.9)	
Extreme	9 (6.4)	8 (22.2)	1 (1.0)	
Discharge disposition, n (%)				**<.001**
Home	147 (90.2)	24 (64.9)	123 (97.6)	
Hospice or home health services	7 (4.3)	5 (13.5)	2 (1.6)	
Expired	5 (3.1)	5 (13.5)	0 (0.0)	
Other	4 (2.5)	3 (8.1)	1 (0.8)	

[a]For continuous variables, independent *t* test was used; for categorical variables, Chi-square test of independence was used. Fisher exact test was used for variables with cell counts of less than 5.

[b]Not available.

[c]*P*<.05.

[d]APR-DRG: all patient refined-diagnosis related group.

Feature Selection

Feature selection was performed to reduce the number of features, and the reduced feature set was fed into each of the classifiers. The sample distribution for each of the six datasets for 3-hour observational periods is given in Table 3. The number of patients varies with the availability of data during each time window.

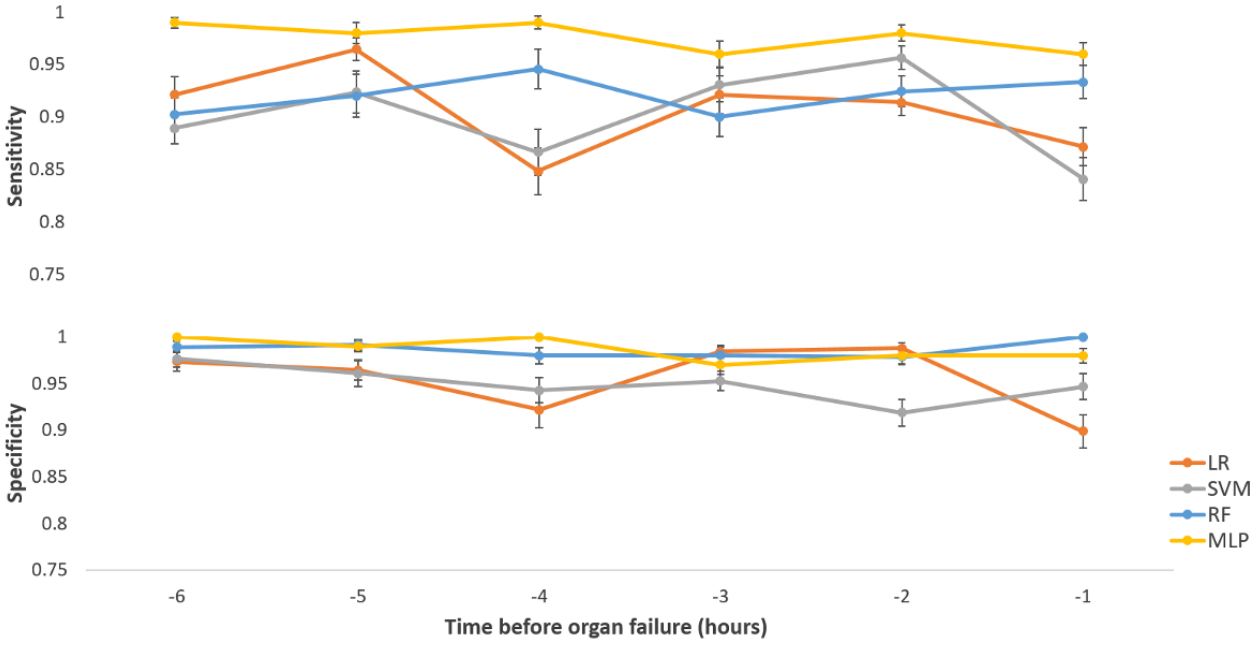
Table 3. Sample distribution and number of features for each dataset using organ failure.

Interval before organ failure onset (hours)	Organ failure events, n	Control events, n
6-9	27	97
5-8	22	90
4-7	22	89
3-6	29	83
2-5	29	88
1-4	29	79

Model Performance

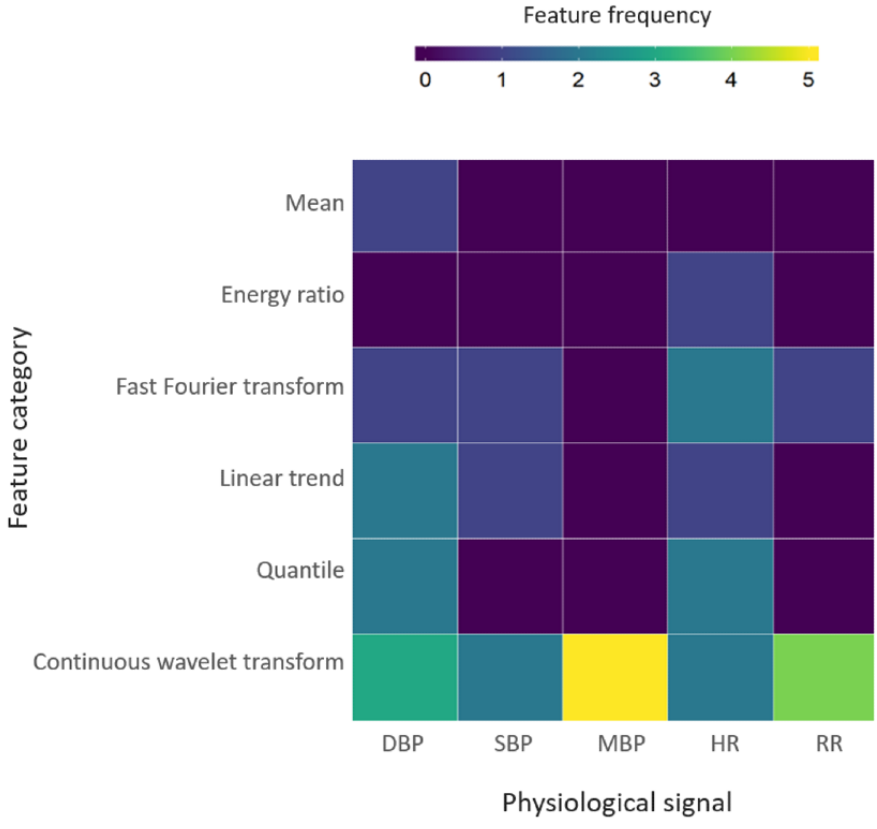
The average sensitivity and specificity from all models for each of the six time periods are given in Figure 2. The MLP model achieved an average sensitivity and specificity of 96% and 98%, respectively, an hour before organ failure (Figure 2 and Multimedia Appendix 3). Among the four classifiers, MLP performed better than SVM, LR, and RF in predicting SCD with organ failure.

Figure 2. Average sensitivity, specificity for support vector machine, random forest, logistic regression, multilayer perceptron, and sickle cell disease models using each of the six 3-hour datasets. LR: logistic regression; SVM: support vector machine; RF: random forest; MLP: multilayer perceptron.



RF classifier identified the continuous wavelet transform generated from MBP time series physiologic variable as the most important feature, followed by continuous wavelet transform feature generated from respiratory rate of the time series data. Figure 3 shows the frequency of the top 30 important features generated from the five physiologic signals. Each box in the heat map represents the frequency of a feature (y-axis) generated from the physiological signal (x-axis). The dark purple color in Figure 3 represents the absence of the feature, whereas the dark yellow color represents the most frequently present feature. The list of features ranked by their importance is given in Multimedia Appendix 4. The description of each feature is presented in Multimedia Appendix 1. The Kruskal-Wallis statistical test found no difference among the features extracted from physiological signals (H statistic=5.029; *P*=.28, possibly reflecting the relatively small sample size.

Figure 3. Features derived from physiologic signals up to six hours before organ failure. DBP: diastolic blood pressure; SBP: systolic blood pressure; MBP: mean blood pressure; HR: heart rate; RR: respiratory rate.



Discussion

Principal Findings

Acute organ failure is a major challenge in people with SCD, especially among adults experiencing an acute disease complication. The ability of predictive algorithms to identify patients at high risk for organ deterioration by using routinely collected physiological data can provide important early warnings of impending physiological deterioration. Such information can aid clinical decision making and may even eventually be useful in guiding early goal-directed therapy. In this retrospective study, we demonstrated that the machine learning–based prediction models could accurately distinguish patients with SCD at risk for developing organ failure up to 6 hours before the onset. The classifiers and the selected physiologic features may facilitate accurate, unbiased SCD diagnosis and effective treatment, ultimately improving prognosis.

The selection of relevant features involved in SCD with organ failure remains a challenge [31,32]. Therefore, we wanted to find a subset of physiologic features that are sufficiently informative to distinguish between patients with SCD at risk of developing organ failure and those who are not at risk. To extract useful information from continuous physiologic data of patients with SCD and to reduce dimensionality, feature-selection algorithms were systematically investigated. As we have demonstrated in the results, selecting smaller subsets of features allowed for the high performance of our classification models. Salient physiomarkers (such as fast Fourier transform, energy, and continuous wavelet transform) derived from the physiological signals, such as blood pressure, HR, and RR, may precede acute organ failure in patients with SCD, as suggested by the results in this study. A shortcoming of machine learning is that these physiomarkers are neither observable by physicians nor readily interpretable; instead, their benefit is primarily toward the early prediction of impending physiologic deterioration, as well as alerting health care providers of that fact. Further research is needed to understand how to use these alerts to guide the personalized care of patients with SCD.

The high-frequency data were captured at 1-min intervals, and we only studied patients who had at least 24 hours of continuous high-frequency physiologic data available before the time of organ failure onset (identified using SOFA criteria). Patients admitted to the ICU with organ failure were excluded, as were patients without a full 24 hour of preceding data. It is also possible that some patients who were too sick may not have been connected to the monitors, which may have introduced the selection bias. There is a need for future research to focus on developing models that rely on less data before organ failure.

Other limitations are also important to mention. First, we developed the machine learning model on a small subset of patients, specific to the Mid-South of the United States, potentially reducing generalizability. Moreover, the data were highly imbalanced, with more non–organ failure cases compared with organ failure cases, making data-driven approaches difficult to implement. Missing data elements identified in the data were a major hindrance for model validation; thus, these may have contributed to poor validation in some of the cross-validation folds. Although we included admission/encounter in the machine learning model building if the intervals between admissions were at least 1 month, a patient may be more likely to have organ failure in the subsequent encounter. With larger patient data, in the future, we can restrict events to a single event per patient.

For the purposes of this particular study, the SOFA scores were used only to classify cases and controls and to determine the time of organ failure onset. The machine learning models to distinguish cases and controls were built using a limited set of continuously streaming physiological data. There is an inherent difference in how SOFA data are collected and used versus how data for machine learning were collected and used. Moreover, the time of SOFA scores is inherently delayed, to some unknown degree, and this leads to noise in any predictive model whose goal is dependent on the timing of an event (such as organ failure). As data to compute SOFA scores may be delayed in being entered into the EMR, our proposed machine learning model could be used as an alternative for timely diagnosis. Therefore, additional data are required to develop a more robust and generalizable model.

Conclusions

In conclusion, we showed, as a proof of principle, that machine learning can accurately predict the development of organ failure in ICU patients with SCD up to 6 hours before onset. This finding is significant because it may optimize the early recognition of serious disease complications and allow for the implementation of early interventions. As future plans, we would like to extend this study to develop a multiclass machine learning classification model to predict the type of organ failure from each of the six organ systems as we collect sufficient data from each organ system.

Acknowledgments

The authors wish to acknowledge Brian Williams, Michael Younker, and Don MacMillan for their assistance in the data collection. The authors would also like to thank the Office of Scientific Writing at the University of Tennessee Health Science Center for copyediting and proofreading the paper. The authors gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp used for this research.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Temporal, frequency, and statistical features.
[XLSX File (Microsoft Excel File), 17 KB - jmir_v22i5e14693_app1.xlsx]

Multimedia Appendix 2
Illustration of feature extraction for cases and controls. Each dashed box represent 3-hour physiological data. For cases, the green and blue arrow represent data from one and two-hour prior to organ failure onset, respectively. For controls, gray arrows represent a random 3-hour physiological data.
[PNG File , 79 KB - jmir_v22i5e14693_app2.png]

Multimedia Appendix 3
Average sensitivity, specificity for multi-layer perceptron (MLP), support vector machine (SVM), random forest (RF), and logistic regression (LR).
[XLSX File (Microsoft Excel File), 11 KB - jmir_v22i5e14693_app3.xlsx]

Multimedia Appendix 4
The list of features ranked by their importance.
[XLSX File (Microsoft Excel File), 10 KB - jmir_v22i5e14693_app4.xlsx]

References

1. Powars DR. Sickle cell anemia and major organ failure. Hemoglobin 1990;14(6):573-598. [doi: 10.3109/03630269009046967] [Medline: 2101835]
2. Thein MS, Igbineweka NE, Thein SL. Sickle cell disease in the older adult. Pathology 2017 Jan;49(1):1-9. [doi: 10.1016/j.pathol.2016.10.002] [Medline: 27914684]
3. Powars DR, Chan LS, Hiti A, Ramicone E, Johnson C. Outcome of sickle cell anemia: a 4-decade observational study of 1056 patients. Medicine (Baltimore) 2005 Nov;84(6):363-376 [FREE Full text] [doi: 10.1097/01.md.0000189089.45003.52] [Medline: 16267411]
4. Johnson AE, Ghassemi MM, Nemati S, Niehaus KE, Clifton DA, Clifford GD. Machine learning and decision support in critical care. Proc IEEE Inst Electr Electron Eng 2016 Feb;104(2):444-466 [FREE Full text] [doi: 10.1109/JPROC.2015.2501978] [Medline: 27765959]
5. Kamaleswaran R, Akbilgic O, Hallman MA, West AN, Davis RL, Shah SH. Applying artificial intelligence to identify physiomarkers predicting severe sepsis in the PICU. Pediatr Crit Care Med 2018 Oct;19(10):e495-e503. [doi: 10.1097/PCC.0000000000001666] [Medline: 30052552]
6. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med 1985 Oct;13(10):818-829. [Medline: 3928249]
7. Marshall JC, Cook DJ, Christou NV, Bernard GR, Sprung CL, Sibbald WJ. Multiple organ dysfunction score: a reliable descriptor of a complex clinical outcome. Crit Care Med 1995 Oct;23(10):1638-1652. [doi: 10.1097/00003246-199510000-00007] [Medline: 7587228]
8. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). J Am Med Assoc 2016 Feb 23;315(8):801-810 [FREE Full text] [doi: 10.1001/jama.2016.0287] [Medline: 26903338]
9. Vincent JL, Moreno R, Takala J, Willatts S, de Mendonça A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. Intensive Care Med 1996 Jul;22(7):707-710. [doi: 10.1007/bf01709751] [Medline: 8844239]
10. van Wyk F, Khojandi A, Mohammed A, Begoli E, Davis RL, Kamaleswaran R. A minimal set of physiomarkers in continuous high frequency data streams predict adult sepsis onset earlier. Int J Med Inform 2019 Feb;122:55-62. [doi: 10.1016/j.ijmedinf.2018.12.002] [Medline: 30623784]
11. Christ M, Braun N, Neuffer J, Kempa-Liehr AW. Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package). Neurocomputing 2018;307:72-77. [doi: 10.1016/j.neucom.2018.03.067]
12. Bishop CM. Pattern Recognition and Machine Learning. Cham: Springer; 2007.
13. Statnikov A, Wang L, Aliferis CF. A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. BMC Bioinformatics 2008 Jul 22;9:319 [FREE Full text] [doi: 10.1186/1471-2105-9-319] [Medline: 18647401]
14. Tiwari AK. Machine learning based approaches for prediction of Parkinson's disease. Mach Learn Appl An Int J 2016;3(2):33-39. [doi: 10.5121/mlaij.2016.3203]
15. Orhan U, Hekim M, Ozer M. EEG signals classification using the K-means clustering and a multilayer perceptron neural network model. Expert Syst Appl 2011;38(10):13475-13481. [doi: 10.1016/j.eswa.2011.04.149]

16. Salem AM, Revett K, El-Dahshan EA. Machine Learning in Electrocardiogram Diagnosis. In: Proceedings of the 2009 International Multiconference on Computer Science and Information Technology. 2009 Presented at: IMCSIT'09; October 12-14, 2009; Mragowo, Poland. [doi: 10.1109/IMCSIT.2009.5352689]
17. Mena L, Gonzalez JA. Symbolic one-class learning from imbalanced datasets: application in medical diagnosis. Int J Artif Intell Tools 2009;18(2):273-309. [doi: 10.1142/s0218213009000135]
18. Magoulas GD, Prentza A. Machine learning in medical applications. In: Paliouras G, Karkaletsis V, Spyropoulos CD, editors. Machine Learning and Its Applications. Berlin, Heidelberg: Springer; 2001:300-307.
19. Itoh N, Hirashiki KI, Terada T, Kikuta M, Ishizuka SI, Koto T, et al. High sensitivity 900-MHz ISM band transceiver. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences 2005 Feb 1;E88-A(2):498-506. [doi: 10.1093/ietfec/e88-a.2.498]
20. Yan H, Jiang Y, Zheng J, Peng C, Li Q. A multilayer perceptron-based medical decision support system for heart disease diagnosis. Expert Syst Appl 2006 Feb;30(2):272-281. [doi: 10.1016/j.eswa.2005.07.022]
21. Antal P, Fannes G, Timmerman D, Moreau Y, de Moor B. Bayesian applications of belief networks and multilayer perceptrons for ovarian tumor classification with rejection. Artif Intell Med 2003 Sep;29(1-2):39-60. [doi: 10.1016/s0933-3657(03)00053-8]
22. Morgan N, Bourlard H. Continuous Speech Recognition Using Multilayer Perceptrons With Hidden Markov Models. In: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing. 1990 Presented at: ICASSP'90; April 3-6, 1990; Albuquerque, NM, USA, USA. [doi: 10.1109/ICASSP.1990.115720]
23. Ren J, Shen X, Lin Z, Mech R, Foran DJ. Personalized Image Aesthetics. In: Proceedings of the 2017 IEEE International Conference on Computer Vision. 2017 Presented at: ICCV'17; October 22-29, 2017; Venice, Italy. [doi: 10.1109/ICCV.2017.76]
24. Agarwal A, Negahban S, Wainwright MJ. Noisy matrix decomposition via convex relaxation: Optimal rates in high dimensions. Ann Statist 2012;40(2):1171-1197. [doi: 10.1214/12-aos1000]
25. Zhang H, Gu C. CiteSeerX. 2006. Support Vector Machines versus Boosting URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.117.343&rep=rep1&type=pdf [accessed 2020-02-25]
26. Kajdanowicz T, Kazienko P. Boosting-based multi-label classification. J Univers Comput Sci 2013;19(4):502-520. [doi: 10.3217/jucs-019-04-0502]
27. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. J Biomed Inform 2002;35(5-6):352-359 [FREE Full text] [doi: 10.1016/s1532-0464(03)00034-0] [Medline: 12968784]
28. Martinkova N, Nová P, Sablina OV, Graphodatsky AS, Zima J. Karyotypic relationships of the Tatra vole (Microtus tatricus). Folia Zool 2004;53(3):279-284 [FREE Full text]
29. Rubin DB. The Bayesian Bootstrap. Ann Statist 1981 Jan;9(1):130-134. [doi: 10.1214/aos/1176345338]
30. Statnikov A, Aliferis CF, Tsamardinos I, Hardin D, Levy S. A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis. Bioinformatics 2005 Mar 1;21(5):631-643. [doi: 10.1093/bioinformatics/bti033] [Medline: 15374862]
31. Yeruva SL, Paul Y, Oneal P, Nouraie M. Renal failure in sickle cell disease: prevalence, predictors of disease, mortality and effect on length of hospital stay. Hemoglobin 2016 Sep;40(5):295-299 [FREE Full text] [doi: 10.1080/03630269.2016.1224766] [Medline: 27643740]
32. Al Khawaja SA, Ateya ZM, Al Hammam RA. Predictors of mortality in adults with Sickle cell disease admitted to intensive care unit in Bahrain. J Crit Care 2017 Dec;42:238-242. [doi: 10.1016/j.jcrc.2017.07.032] [Medline: 28797896]

Abbreviations

DBP: diastolic blood pressure
EMR: electronic medical record
HR: heart rate
ICU: intensive care unit
LR: logistic regression
MBP: mean blood pressure
MLP: multilayer perceptron
RF: random forest
RR: respiratory rate
SBP: systolic blood pressure
SCD: sickle cell disease
SOFA: Sequential Organ Failure Assessment
SVM: support vector machine
t_{onset}: event time

Edited by M Focsa; submitted 13.05.19; peer-reviewed by S Creary, S Curtis, C Gao; comments to author 02.06.19; revised version received 18.08.19; accepted 28.01.20; published 13.05.20.

Please cite as:
Mohammed A, Podila PSB, Davis RL, Ataga KI, Hankins JS, Kamaleswaran R
Using Machine Learning to Predict Early Onset Acute Organ Failure in Critically Ill Intensive Care Unit Patients With Sickle Cell Disease: Retrospective Study
J Med Internet Res 2020;22(5):e14693
URL: https://www.jmir.org/2020/5/e14693
doi:10.2196/14693
PMID:32401216

Original Paper

Association Between Web-Based Physician Ratings and Physician Disciplinary Convictions: Retrospective Observational Study

Jessica Janine Liu[1,2], MSc, MD; Hanna R Goldberg[3], MSc, MD; Eric JM Lentz[4], BA; John Justin Matelski[2], MSc; Asim Alam[5,6], MD; Chaim M Bell[1,7], MD, PhD

[1]Division of General Internal Medicine, Department of Medicine, University of Toronto, Toronto, ON, Canada

[2]Department of Medicine, University Health Network, Toronto, ON, Canada

[3]Department of Obstetrics and Gynecology, University of Toronto, Toronto, ON, Canada

[4]Faculty of Medicine, McMaster University, Hamilton, ON, Canada

[5]Department of Anesthesia and Surgery, North York General Hospital, Toronto, ON, Canada

[6]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada

[7]Sinai Health System, Toronto, ON, Canada

Corresponding Author:
Jessica Janine Liu, MSc, MD
Division of General Internal Medicine
Department of Medicine
University of Toronto
Toronto General Hospital
200 Elizabeth Street, Room 14EN213
Toronto, ON, M5G 2C4
Canada
Phone: 1 4163403111
Fax: 1 4165955826
Email: jessica.liu@uhn.ca

Abstract

Background: Physician rating websites are commonly used by the public, yet the relationship between web-based physician ratings and health care quality is not well understood.

Objective: The objective of our study was to use physician disciplinary convictions as an extreme marker for poor physician quality and to investigate whether disciplined physicians have lower ratings than nondisciplined matched controls.

Methods: This was a retrospective national observational study of all disciplined physicians in Canada (751 physicians, 2000 to 2013). We searched ratings (2005-2015) from the country's leading online physician rating website for this group, and for 751 matched controls according to gender, specialty, practice years, and location. We compared overall ratings (out of a score of 5) as well as mean ratings by the type of misconduct. We also compared ratings for each type of misconduct and punishment.

Results: There were 62.7% (471/751) of convicted and disciplined physicians (cases) with web-based ratings and 64.6% (485/751) of nondisciplined physicians (controls) with ratings. Of 312 matched case-control pairs, disciplined physicians were rated lower than controls overall (3.62 vs 4.00; $P<.001$). Disciplined physicians had lower ratings for all types of misconduct and punishment—except for physicians disciplined for sexual offenses (n=90 pairs; 3.83 vs 3.86; $P=.81$). Sexual misconduct was the only category in which mean ratings for physicians were higher than those for other disciplined physicians (3.63 vs 3.35; $P=.003$)

Conclusions: Physicians convicted for disciplinary misconduct generally had lower web-based ratings. Physicians convicted of sexual misconduct did not have lower ratings and were rated higher than other disciplined physicians. These findings may have future implications for the identification of physicians providing poor-quality care.

(J Med Internet Res 2020;22(5):e16708) doi:10.2196/16708

KEYWORDS

quality improvement; patient satisfaction; patient-centered care

Introduction

Background

The ability of patients to accurately evaluate health care quality is not well understood. Although some studies demonstrate an association between greater patient satisfaction and quality of care, others show either no relationship or even poorer outcomes with increased patient satisfaction [1-12]. Over the last decade, with the advent of physician rating websites such as healthgrades.com, ratemds.com, and vitals.com, a novel source of patient satisfaction data has emerged. Such websites have become popular forums for patients to evaluate and publicly share their health care experience.

Previous studies have focused on the awareness of and frequency of ratings for specific medical specialties on rating websites in Canada, the United States, China, and Germany [11-29]. The recent focus has been to correlate web-based ratings with quality outcomes or surrogates such as postoperative mortality—with variable findings [13,30-35]. Web-based physician ratings represent a novel, unsolicited data source of the patient experience with health care providers that is unique from more traditional satisfaction measures, such as solicited surveys.

Physician misconduct can be considered a reflection of poor quality care. Physicians are investigated, convicted, and disciplined by their professional associations for activities such as unprofessional behavior, sexual misconduct, failure to meet standards of care, fraud, abuse of drugs and alcohol, and negligence. Resultant penalties range from fines and mandatory education to license suspension and revocation. Although disciplinary proceedings are publicly posted by each province's physician regulatory college, at the time of a clinical encounter, patients are often unaware of a physician's disciplinary history.

Objectives

We were interested in whether or not physicians who have been convicted and punished for misconduct are rated differently than nondisciplined physician controls. We hypothesized that, for many types of misconduct, patients would accurately recognize poor-quality physicians and felt that, overall, disciplined physicians would have lower web-based ratings than controls. We also sought to determine whether ratings were consistently lower across all types of misconduct, and we hypothesized that associations between ratings and discipline would differ depending on the type of misconduct.

Methods

Physician Databases: Disciplined Physicians

This retrospective cohort study reviewed publicly available information on physician disciplinary proceedings published by Canadian provincial and territorial physician regulatory colleges. A database from January 2000 to December 2013 was compiled (described previously) [36-40]. We collected demographic information for every disciplined physician in the country, including gender, license type (independent vs educational), medical school (ie, North American trained versus international medical graduate [IMG]), year of graduation, and specialty. We collected information on types of misconduct and resultant penalties that were determined by the provincial colleges. Misconduct was categorized into (1) inappropriate prescribing, (2) criminal conviction, (3) fraudulent behavior or prevarication, (4) misconduct secondary to mental illness, (5) self-use of drugs or alcohol, (6) sexual misconduct, (7) practice below standard of care, (8) unprofessional conduct, (9) unlicensed activity, (10) miscellaneous findings (ie, improper maintenance of medical records and confidentiality breaches), and (11) unclear. Punishments included (1) license revocation, (2) voluntary license surrender, (3) suspension, (4) license restriction, (5) mandated retraining, education or assessment, (6) mandated participation in psychological counseling or addiction rehabilitation, (7) formal reprimand, (8) fine or cost repayment, and (9) other [36-42].

Cases and Controls

We employed a nested case-control design and matched each disciplined physician (*cases*) with a nondisciplined counterpart (*controls*) according to specialty, gender, town of listed practice, and years in medical practice (within 5 years). We developed a group of nondisciplined physician controls by searching provincial physician regulatory college websites for each disciplined physician and narrowing our search terms by the abovementioned criteria. In certain instances (ie, 2 provinces), if after controlling for the 4 matching criteria, multiple physician matches were possible, a physician was chosen at random. In total, 751 disciplined physicians were matched with 751 nondisciplined controls. A nondisciplined control was found for every disciplined physician (Table 1).

Table 1. Characteristics of cases (751 disciplined physicians; 2000-2013) and controls (751 nondisciplined physicians, matched for gender, years since graduation, and city of practice [where possible]).

Physician characteristic	Cases (disciplined physicians; N=751)			Controls (nondisciplined controls; N=751)		
	Rated matched (n=312)	Rated unmatched (n=159)	Unrated (n=280)	Rated matched (n=312)	Rated unmatched (n=173)	Unrated (n=266)
Sex, n (%)						
Female	28 (9.0)	14 (8.8)	24 (8.6)	28 (9)	15 (8.7)	23 (9.6)
Male	284 (91)	145 (91.2)	256 (91.4)	284 (91)	158 (91.3)	243 (91.4)
Years in practice since graduation, mean (SD)	27.9 (10.4)	29.9 (11.7)	30.7 (12.0)	27.3 (10.9)	27.2 (11.4)	29.8 (11.9)
Specialty, n (%)						
Family medicine	180 (57.7)	94 (59.1)	167 (59.6)	180 (57.7)	113 (65.3)	152 (57.1)
Internal medicine	7 (2.2)	2 (1.3)	14 (5.0)	6 (1.9)	8 (4.6)	10 (3.8)
Obstetrics	24 (7.7)	2 (1.9)	5 (1.8)	24 (7.7)	5 (2.9)	3 (1.1)
Pediatrics	6 (1.9)	5 (3.1)	2 (0.7)	6 (1.9)	1 (0.6)	6 (2.3)
Psychiatry	23 (7.4)	23 (14.5)	46 (16.4)	23 (7.4)	19 (11.0)	47 (17.7)
Radiology	0 (0)	1 (0.6)	3 (1.1)	1 (0.3)	2 (1.2)	2 (0.8)
Surgery	23 (7.4)	7 (4.4)	25 (8.9)	24 (7.7)	20 (11.6)	13 (4.9)
Other	49 (15.7)	20 (12.5)	12 (4.3)	48 (15.4)	4 (2.3)	22 (8.2)
Medical school, n (%)						
International medical graduate[a]	91 (29.2)	50 (31.4)	108 (38.6)	104 (33.3)	59 (34.1)	113 (42.4)
North American graduate	221 (70.8)	109 (68.6)	172 (61.4)	208(66.7)	14 (65.9)	153 (57.5)
Number of ratings, mean (SD)	19.6 (13.0)	15.0 (11.82)	N/A[b]	15.5 (11.5)	14.9 (10.5)	N/A
Overall rating, mean (SD)	3.62 (0.82)	3.42 (0.98)	N/A	4.00 (0.75)	3.91 (0.82)	N/A

[a]International medical graduate denotes physicians who graduated from a non-Canadian or non-US medical school.

[b]N/A: not applicable.

Physician Ratings Data

RateMDs.com is a publicly accessible physician rating website founded in the United States in 2004. Since its launch in Canada (2005), it is the country's leading physician rating website and one of the most popular physician rating websites in North America [41,43]. As of 2013, RateMDs.com included more than 640,000 ratings of over 57,000 unique physicians in Canada [29]. No registration or subscription is required to view or submit a rating, and there are no monetary reimbursement or other incentives to rate a physician. Physicians are rated on staff (typically front office staff), punctuality, helpfulness, and knowledge, on a scale of 1 to 5 (1=*terrible*, 2=*poor*, 3=*okay*, 4=*good*, and 5=*excellent*). Raters may provide text comments if desired. It must be noted that RateMDs.com does not provide disciplinary information. We reviewed all disciplined and nondisciplined control physicians on this website and recorded rating scores. Data collection took place between approximately May 2014 and September 2014, with data cleaning and quality control performed by a second party in July 2015.

Creation of the Dataset

We paired 751 disciplined physicians with 751 nondisciplined matched controls and collected information from rateMDs.com for each physician. As not all physicians were rated, this resulted in 4 groups: disciplined rated *cases*, disciplined unrated, control rated, and control unrated. When considering *pairs* of cases and controls who both had web-based ratings, our dataset included 312 physician pairs (Figure 1). We used only matched pairs for analysis and performed analyses when there were more than 50 case-control matched pairs. We also grouped disciplined physicians according to types of misconduct and punishments. The number of matched pairs available for testing varied from 2 to 254 pairs available for comparison (Table 2).

Figure 1. Flow chart for developing matched case (disciplined) and control (nondisciplined) groups of physicians, according to the presence of at least one web-based rating.

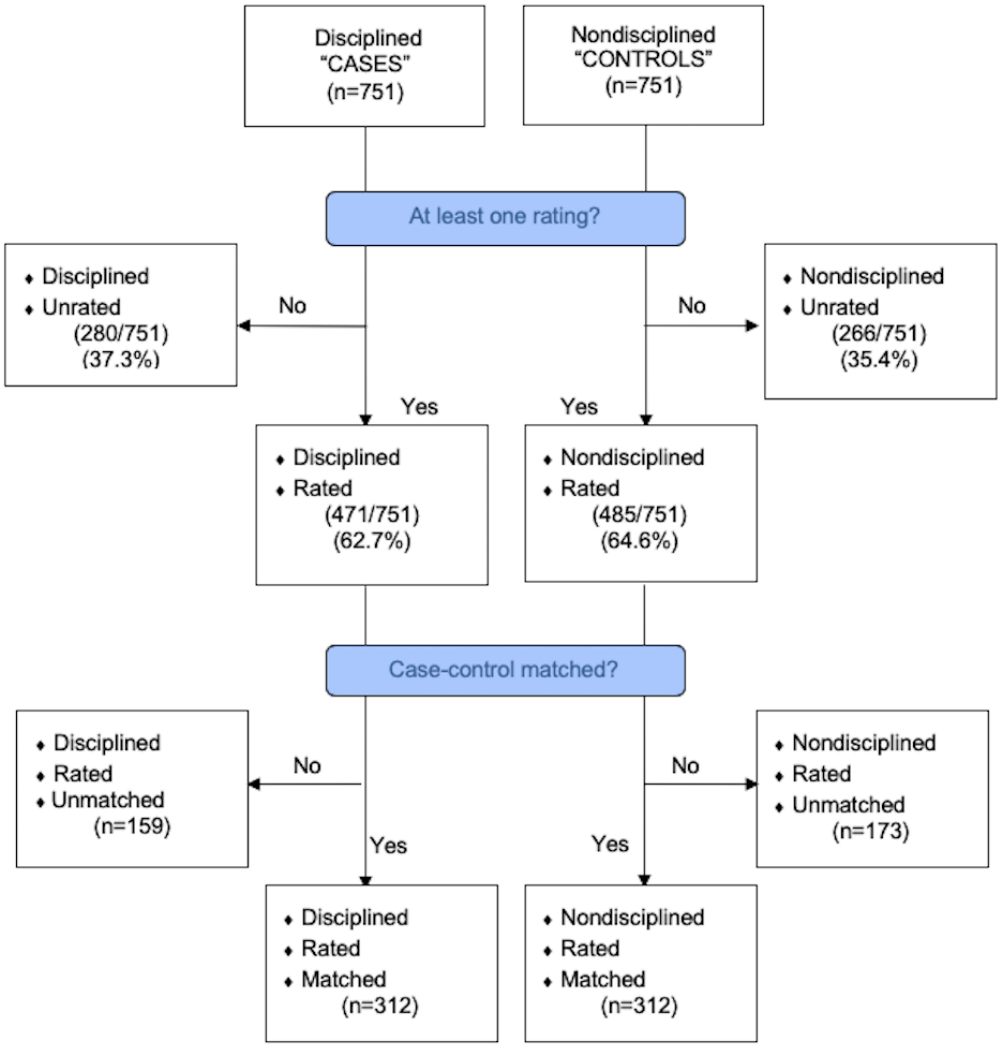


Table 2. Mean web-based physician ratings of disciplined physicians (2000-2013) and nondisciplined controls (2000-2013) by the type of misconduct and punishment.

Description	Case-control pairs	Disciplined physicians (cases)		Nondisciplined matched controls		Statistical analysis
		Mean rating	95% CI	Mean rating	95% CI	Wald *P* value
Overall	312	3.62	3.53-3.71	4.00	3.91-4.08	<.001
Physicians with more than one conviction (repeat offenders)	44	3.44	3.21-3.85	4.00	3.67-4.33	<.01
Misconduct						
Standard of care breach	113	3.57	3.42-3.72	3.98	3.98-3.84	<.001
Inappropriate prescribing	55	3.50	3.29-3.71	3.97	3.79-4.16	.001
Unlicensed activities	35	3.42	3.18-3.66	4.35	4.16-4.54	N/A[a]
Sexual misconduct	90	3.83	3.67-3.99	3.86	3.68-4.04	.81
Mental illness	2	3.89	3.61-4.17	3.26	3.05-3.47	N/A
Drugs/alcohol abuse	8	4.25	3.94-4.55	3.74	3.29-4.20	N/A
Fraudulent behavior	39	3.44	3.18-3.70	4.11	3.89-4.32	N/A
Conviction of a crime	13	3.59	3.13-4.04	4.27	3.95-4.59	N/A
Unprofessional behavior	62	3.43	3.22-3.64	4.03	3.86-4.21	<.001
Miscellaneous	79	3.45	3.29-3.89	3.99	3.82-4.17	<.001
Punishment						
License revocation	25	3.59	3.29-3.89	4.21	3.91-4.51	N/A
License surrender	8	3.98	3.41-4.55	3.38	2.67-4.09	N/A
License suspension	191	3.71	3.60-3.83	3.99	3.88-4.10	<.001
Restriction	115	3.69	3.55-3.83	3.91	3.78-4.05	.03
Mandatory retraining	120	3.61	3.48-3.73	3.98	3.85-4.12	<.001
Counseling	43	3.58	3.35-3.80	3.78	3.42-4.03	N/A
Formal reprimand	144	3.58	3.45-3.71	4.02	3.89-4.14	<.001
Other punishment	34	3.52	3.26-3.78	4.04	3.85-4.22	N/A
Fine	254	3.60	3.50-3.70	4.02	3.93-4.11	<.001

[a]N/A: not applicable (when there were less than 50 case-control pairs available for comparison, analysis was not completed).

Statistical Analysis

Analysis 1: Comparison of Disciplined Versus Nondisciplined Physicians (Matched Analysis)

To compare ratings between disciplined and nondisciplined physicians, we computed an overall average rating for each physician using the mean of the available rating categories, then calculating an overall weighted mean. Generalized estimating equations (GEEs) were used to estimate the average rating by group (disciplined vs nondisciplined), and GEES were used for each type of misconduct or penalty. GEEs were selected to account for the matched study design. We felt it was appropriate to select GEEs over nonparametric testing, given that there were sufficiently large (eg, 451) distinct average ratings, and therefore, it could treat ordinal data similar to continuous data. This analysis allowed us to report 95% CIs for the estimated group means and provide a sense of the precision of the estimates in addition to significance testing. We reported the estimated mean ratings by group, 95% CIs for these estimates, and Wald *P* value against the null hypothesis (no group difference). An α of .05 was used as the threshold for statistical significance. Analyses were performed using the *geepack* package in R version 3.0.3 (R Foundation).

Analysis 2: Comparison of Physicians Disciplined for a Specific Type of Misconduct/Punishment Versus the Rest of the Disciplined Physicians Cohort

Recognizing that the severity of physician misconduct and punishment is variable (eg, ranging from substandard recordkeeping to more egregious offenses such as sexual misconduct), we compared physician ratings for specific disciplinary offenses with those of the *at large* disciplined physicians cohort. Mixed effects models were used for analyses of ratings among disciplined physicians, considering each physician's overall average web-based rating and category-specific ratings as outcomes. The presence of each type of misconduct/punishment in a physician's discipline record was used as a binary predictor. Gender, year of offense,

province, professional years, and IMG status were included as fixed effects, and physician specialty was included as a random effect. The estimates reflect the mean centering of the year of offense and professional relative to the rest of the disciplined cohort. We report the estimated mean ratings by group, 95% CIs for these estimates, and Wald *P* value against the null hypothesis (no group difference). An α of .05 is used as the threshold for statistical significance. Analyses were performed using the *nlme* package in R version 3.0.3.

Sensitivity Analysis

To assess the degree to which physicians with a low overall number of ratings (ie, <5 or <10 ratings) influenced our overall results, we performed additional testing on both analyses 1 and 2 by excluding instances in which physicians had (1) less than 5 overall ratings and (b) less than 10 overall ratings.

Results

Disciplined Physicians Versus Nondisciplined Physicians: Matched Analysis

We paired 751 disciplined physicians with 751 nondisciplined matched controls. Of the 751 disciplined physicians, 37.3% (280/751) did not have any web-based ratings, whereas 62.7% (471/751) had at least one rating. Of the 751 nondisciplined physician controls, 64.6% (485/751) had at least one rating, whereas 35.4% (266/751) were not rated online. When comparing rated, but unmatched, physicians, 21.1% (159/751) were disciplined, rated, but unmatched compared with 23.0% (173/751) nondisciplined, rated, but unmatched. When considering *pairs* of cases and controls who both had ratings, our dataset included 312 physician pairs (Figure 1). When we grouped disciplined physicians according to the types of misconduct and punishments, the number of matched pairs available varied, ranging from 2 to 254 available pairs (Table 2).

When we compared the 312 pairs of convicted and disciplined physicians with nondisciplined controls, disciplined physicians were rated lower than nondisciplined physicians for all offenses and punishments (mean rating 3.62, SD 0.82 vs mean rating 4.00, SD 0.75; *P*<.001). When comparing rated, but unmatched, physicians, disciplined unmatched physicians had even lower ratings than nondisciplined unmatched physicians (mean 3.42, SD 0.98 vs mean 3.91, SD 0.82). As 12.5% (94/751) of our disciplined physicians cohort had more than one disciplinary conviction during our study period, we also looked at this group of *repeat offenders*. Of the 94 disciplined physicians who were repeat offenders, approximately half were available for case-control analysis, as 44 disciplined physicians were appropriately matched to a case-control where both groups had ratings. Disciplined repeat offenders had mean ratings that were also lower than controls (mean 3.44, SD 4.09 vs mean 4.00, SD 0.81; *P*<.01).

The mean rating for disciplined physicians was lower than that for nondisciplined physician–matched controls for the following types of misconduct and punishment: standard of care breach (113 pairs; 3.57 vs 3.98; *P*<.001), inappropriate prescribing (55 pairs; 3.50 vs 3.97; *P*<.001), unprofessional behavior (62 pairs; 3.43 vs 4.03; *P*<.001), miscellaneous/unclear (79 pairs; 3.45 vs 3.99; *P*<.001), license suspension (191 pairs; 3.71 vs 3.99; *P*<.001), license restriction (115 pairs; 3.69 vs 3.91; *P*=.027), mandatory retraining (120 pairs; 3.61 vs 3.98; *P*<.001), formal reprimand (144 pairs; 3.58 vs 4.02; *P*<.001), and fine (254 pairs; 3.60 vs 4.02; *P*<.001; Table 1). No significant differences were detected for physicians who were disciplined for sexual offenses, compared with nondisciplined matched controls (n=90 physician pairs; 3.83 vs 3.86; *P*=.81; Table 2).

Comparison of Specific Type of Misconduct/Punishment Versus the Rest of the Disciplined Physicians Cohort

Sexual misconduct was the only category of misconduct in which mean ratings for this group of physicians were higher than those for other disciplined physicians. Moreover, 62.7% (471/751) disciplined physicians who were rated online, 219 were disciplined for sexual misconduct. The overall mean rating of physicians disciplined for sexual misconduct was higher than that of all other disciplined physicians (3.63; 95% CI 3.17-4.08 vs 3.35, 95% CI 2.91-3.80; *P*=.003). This overall effect was consistent and significant across all 4 rating subcategories (*staff*: 3.93 vs 3.71; *P*=.023; *punctuality*: 3.60 vs 3.36; *P*=.011; *helpfulness*: 3.83 vs 3.47; *P*<.001; and *knowledge*: 4.02 vs 3.66; *P*<.001; Multimedia Appendix 1). Physicians disciplined for fraudulent behavior and miscellaneous had *lower* overall ratings when compared with other disciplined physicians (fraudulent behavior: 3.15 vs 3.44; *P*=.01 and miscellaneous: 3.31 vs 2.51; *P*=.04). For punishments, suspension was the only type of punishment in which this group of disciplined physicians was rated higher than all other disciplined physicians (3.54 vs 3.33; *P*=.023); however, this result did not remain robust when physicians with less than 10 ratings were excluded from our sensitivity analysis. For all other types of misconduct and punishments, no overall mean rating differences existed compared with all other disciplined physicians (Multimedia Appendix 1).

Sensitivity Analysis

Of 312 cases and 312 controls, 48 case physicians and 68 control physicians had less than 5 ratings. Similarly, 84 case physicians and 117 control physicians had less than 10 ratings. To assess whether such ratings influenced our main results, we performed sensitivity analyses by excluding cases in which physicians had (1) less than 5 and (2) less than 10 ratings. Our main results remained robust (Multimedia Appendices 2 and 3). When we excluded physicians with few ratings, our finding that disciplined physicians had lower overall mean ratings did not change (<5 ratings: 3.61 vs 4.01; *P*<.001 and <10 ratings: 3.52 vs 4.01; *P*<.001). When broken down by type of misconduct and punishment, results also remained robust—that is, disciplined physicians had lower ratings than nondisciplined case-controls, with the exception of sexual misconduct (<5 ratings and sexual misconduct: 3.89 vs 3.94; *P*=.69 and <10 ratings and sexual misconduct: 3.79 vs 3.92; *P*=0.36).

Similarly, when comparing physicians disciplined for types of misconduct with all other disciplined physicians, all results remained robust, with 2 minor exceptions. Ratings for physicians whose licenses were suspended no longer differed from all other

disciplined physicians, nor did the ratings for physicians who were punished with a formal reprimand (Multimedia Appendix 3). All other results remained consistent after sensitivity analyses.

Discussion

Principal Findings

Our study used a national dataset of all disciplined physicians and collected their available online ratings from rateMDs.com over a 10-year period. Of over 750 matched physician pairs, 63.6% (956/1502) physicians are rated online. For most types of misconduct, disciplined physicians are rated lower than nondisciplined controls. However, physicians disciplined for sexual misconduct were not rated differently than controls and, in fact, were rated *higher* when compared with all other disciplined physicians, a directional relationship that was not found with any other type of misconduct.

Comparison With Prior Work

Our results are in general agreement with other studies that show that physicians are, overall, rated positively [13,20,21,29]. Our findings are also consistent with data showing lower online ratings for physicians on probation for many types of misconduct, but not sexual offenses [34]. There may be something unique about physicians who commit sexual misconduct that distinguishes them from other convicted physicians, at least with respect to online ratings.

We found that online raters discerned a difference between disciplined and nondisciplined physicians with respect to online ratings overall; however, interestingly, sexual misconduct was the only category in which this effect was not seen. Furthermore, we found that physicians who were disciplined for sexual misconduct are rated *more favorably* than the rest of the disciplined physician cohort. Again, sexual misconduct was the only category of misconduct in which mean ratings were higher than all other disciplined physicians.

Our findings related to sexual offense convictions are consistent with previous findings. Only a handful of studies have compared sexual offender physicians with other physicians; however, it has been reported that some antisocial personality traits were unique to psychiatrists who were subsequently convicted of sexual boundary violations and that these characteristics were identifiable early in training [44-47].

This study adds to the body of literature on online physician ratings and extends current knowledge to include extremes of poor quality (ie, physician disciplinary convictions). This is the first study to combine 2 large, comprehensive national databases of physician discipline and web-based physician ratings over a 13-year period, using a rigorous matched control approach. We highlight the heterogeneity of disciplined physicians as a group and are among the first to identify this finding in physician sexual offenders. Although the majority of low-rated physicians are not disciplined and they are not sexual offenders, we feel that the potential for patient harm is sufficient enough in such cases to warrant further investigation of this group of disciplined physicians. Future studies could focus on predicting or developing interventions to prevent patient harm.

Limitations

We recognize several limitations. First, our study assumes that disciplined physicians, as a group, are poor-quality physicians. Although not perfectly synonymous, these physicians have been convicted by their professional colleges for conduct that is substandard, inappropriate, or morally not in line with professional standards. As such, this is an excellent surrogate for poor quality. Second, we cannot exclude that publicly posted ratings may, themselves, influence future ratings. Although we considered censoring ratings after a particular disciplinary proceeding became a public record, we felt a time-based analysis would decrease the number of ratings in our analysis, with no clear added benefit against potential bias. Moreover, the uncertainty of whether the rater had advance knowledge of the physician would remain, as it would be difficult to ascertain whether raters were influenced by other sources (eg, popular media attention). Interestingly, we found that physicians who were disciplined for sexual misconduct (the misconduct category frequently reported in the media) were rated no differently than controls. In fact, they were rated *higher* when compared with the rest of disciplined physicians, making us more likely to accept our findings. Although occasionally there was a mention of misconduct in the comments, we estimated this to reflect a small proportion (ie, <5%) of all comments. Moreover, we would argue that for our research question, timing may be less relevant, that is, a physician disciplined in 2000 and reviewed in 2005 versus a physician who was reviewed in 2000 and disciplined in 2005 are both relevant enough to merit consideration.

Third, although we used a stringent matching process, in smaller centers, it was not possible to match by subspecialty for 5 physicians. In this case, we matched as closely as possible (ie, we matched surgeons with another surgeon rather than, eg, a psychiatrist). This represented less than 1% of cases. Fourth, as not all physicians are rated on websites, data may not be generalizable. However, 63.6% (956/1502) of physicians had an online presence, which is much higher than in previous studies [13,15,31], and when we analyzed data from disciplined, unmatched physicians, overall demographics and mean ratings did not substantially differ. In fact, ratings of unmatched, disciplined physicians were lower than unmatched, undisciplined physicians. We also considered external validity concerns in potential comparisons between the 60% of physicians who are rated online versus those who are unrated. However, because our physician control group was hand selected to resemble the disciplined physician group, and not representative of the general population, such comparisons would not be particularly useful; therefore, we specifically refrained from making such direct comparisons between rated and unrated physicians. Finally, rating website users may be different with respect to access to a computer and inclination to post online ratings. However, this is an issue germane to all online ratings. Taken together, we feel that these limitations would not significantly alter our conclusions.

Conclusions

Disciplined physicians are rated lower than control physicians by those who rate their physicians online, in keeping with the

hypothesis that patients can accurately appraise health care quality. However, any ability to ascertain quality becomes more difficult for physicians disciplined for sexual misconduct. Our findings suggest that this group of physicians deserves further investigation to better understand why they would be rated more favorably than all other disciplined physicians. Our research may have implications for the identification of at-risk physicians to develop interventions before patient harm can occur [48].

Authors' Contributions

All the authors (JL, HG, AA, EL, JM, and CB) made substantial contributions to the conception or design of the work or the acquisition, analysis, or interpretation of the work; contributed to drafting the work or revising it critically for important intellectual content; gave final approval for the version to be published; and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Physicians disciplined for specific episodes of misconduct, as compared with other disciplined physicians (2000-20013), broken down by website rating category.
[DOCX File , 25 KB - jmir_v22i5e16708_app1.docx]

Multimedia Appendix 2
Sensitivity Analysis. Mean online ratings (with 95% confidence intervals) for cases versus controls, using inclusion cut-offs of at least 1 online rating (original analysis), at least 5 ratings or at least 10 ratings per physician.
[DOCX File , 17 KB - jmir_v22i5e16708_app2.docx]

Multimedia Appendix 3
Sensitivity Analysis. Mean online ratings for physicians disciplined for specific types of misconduct, as compared with all other disciplined physicians, using inclusion cut-offs of at least 1 online rating, at least 5 ratings or at least 10 ratings per physician.
[DOCX File , 15 KB - jmir_v22i5e16708_app3.docx]

References

1. Stelfox HT, Gandhi TK, Orav EJ, Gustafson ML. The relation of patient satisfaction with complaints against physicians and malpractice lawsuits. Am J Med 2005 Oct;118(10):1126-1133. [doi: 10.1016/j.amjmed.2005.01.060] [Medline: 16194644]
2. Hickson GB, Federspiel CF, Pichert JW, Miller CS, Gauld-Jaeger J, Bost P. Patient complaints and malpractice risk. J Am Med Assoc 2002 Jun 12;287(22):2951-2957. [doi: 10.1001/jama.287.22.2951] [Medline: 12052124]
3. Boulding W, Glickman SW, Manary MP, Schulman KA, Staelin R. Relationship between patient satisfaction with inpatient care and hospital readmission within 30 days. Am J Manag Care 2011 Jan;17(1):41-48 [FREE Full text] [Medline: 21348567]
4. Glickman SW, Boulding W, Manary M, Staelin R, Roe MT, Wolosin RJ, et al. Patient satisfaction and its relationship with clinical quality and inpatient mortality in acute myocardial infarction. Circ Cardiovasc Qual Outcomes 2010 Mar;3(2):188-195. [doi: 10.1161/CIRCOUTCOMES.109.900597] [Medline: 20179265]
5. Sacks GD, Lawson EH, Dawes AJ, Russell MM, Maggard-Gibbons M, Zingmond DS, et al. Relationship between hospital performance on a patient satisfaction survey and surgical quality. JAMA Surg 2015 Sep;150(9):858-864. [doi: 10.1001/jamasurg.2015.1108] [Medline: 26108091]
6. Kane RL, Maciejewski M, Finch M. The relationship of patient satisfaction with care and clinical outcomes. Med Care 1997 Jul;35(7):714-730. [doi: 10.1097/00005650-199707000-00005] [Medline: 9219498]
7. Kennedy GD, Tevis SE, Kent KC. Is there a relationship between patient satisfaction and favorable outcomes? Ann Surg 2014 Oct;260(4):592-8; discussion 598 [FREE Full text] [doi: 10.1097/SLA.0000000000000932] [Medline: 25203875]
8. Chang JT, Hays RD, Shekelle PG, MacLean CH, Solomon DH, Reuben DB, et al. Patients' global ratings of their health care are not associated with the technical quality of their care. Ann Intern Med 2006 May 2;144(9):665-672. [doi: 10.7326/0003-4819-144-9-200605020-00010] [Medline: 16670136]
9. Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ, Lucas FL, Pinder ?. The implications of regional variations in Medicare spending. Part 2: health outcomes and satisfaction with care. Ann Intern Med 2003 Feb 18;138(4):288-298. [doi: 10.7326/0003-4819-138-4-200302180-00007] [Medline: 12585826]
10. Fenton JJ, Jerant AF, Bertakis KD, Franks P. The cost of satisfaction: a national study of patient satisfaction, health care utilization, expenditures, and mortality. Arch Intern Med 2012 Mar 12;172(5):405-411. [doi: 10.1001/archinternmed.2011.1662] [Medline: 22331982]

11. Okike K, Peter-Bibb TK, Xie KC, Okike ON. Association between physician online rating and quality of care. J Med Internet Res 2016 Dec 13;18(12):e324 [FREE Full text] [doi: 10.2196/jmir.6612] [Medline: 27965191]
12. Liu JJ, Matelski J, Cram P, Urbach DR, Bell CM. Association between online physician ratings and cardiac surgery mortality. Circ Cardiovasc Qual Outcomes 2016 Nov;9(6):788-791. [doi: 10.1161/CIRCOUTCOMES.116.003016] [Medline: 27803091]
13. Gao GG, McCullough JS, Agarwal R, Jha AK. A changing landscape of physician quality reporting: analysis of patients' online ratings of their physicians over a 5-year period. J Med Internet Res 2012 Feb 24;14(1):e38 [FREE Full text] [doi: 10.2196/jmir.2003] [Medline: 22366336]
14. Emmert M, Meier F. An analysis of online evaluations on a physician rating website: evidence from a German public reporting instrument. J Med Internet Res 2013 Aug 6;15(8):e157 [FREE Full text] [doi: 10.2196/jmir.2655] [Medline: 23919987]
15. Emmert M, Sander U, Pisch F. Eight questions about physician-rating websites: a systematic review. J Med Internet Res 2013 Feb 1;15(2):e24 [FREE Full text] [doi: 10.2196/jmir.2360] [Medline: 23372115]
16. Hao H. The development of online doctor reviews in China: an analysis of the largest online doctor review website in China. J Med Internet Res 2015 Jun 1;17(6):e134 [FREE Full text] [doi: 10.2196/jmir.4365] [Medline: 26032933]
17. Merrell JG, Levy BH, Johnson DA. Patient assessments and online ratings of quality care: a 'wake-up call' for providers. Am J Gastroenterol 2013 Nov;108(11):1676-1685. [doi: 10.1038/ajg.2013.112] [Medline: 24192941]
18. Terlutter R, Bidmon S, Röttl J. Who uses physician-rating websites? Differences in sociodemographic variables, psychographic variables, and health status of users and nonusers of physician-rating websites. J Med Internet Res 2014 Mar 31;16(3):e97 [FREE Full text] [doi: 10.2196/jmir.3145] [Medline: 24686918]
19. Lagu T, Hannon NS, Rothberg MB, Lindenauer PK. Patients' evaluations of health care providers in the era of social networking: an analysis of physician-rating websites. J Gen Intern Med 2010 Sep;25(9):942-946 [FREE Full text] [doi: 10.1007/s11606-010-1383-0] [Medline: 20464523]
20. Kadry B, Chu LF, Kadry B, Gammas D, Macario A. Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. J Med Internet Res 2011 Nov 16;13(4):e95 [FREE Full text] [doi: 10.2196/jmir.1960] [Medline: 22088924]
21. Emmert M, Meier F, Heider A, Dürr C, Sander U. What do patients say about their physicians? an analysis of 3000 narrative comments posted on a German physician rating website. Health Policy 2014 Oct;118(1):66-73. [doi: 10.1016/j.healthpol.2014.04.015] [Medline: 24836021]
22. Frost C, Mesfin A. Online reviews of orthopedic surgeons: an emerging trend. Orthopedics 2015 Apr;38(4):e257-e262. [doi: 10.3928/01477447-20150402-52] [Medline: 25901617]
23. Ellimoottil C, Hart A, Greco K, Quek ML, Farooq A. Online reviews of 500 urologists. J Urol 2013 Jun;189(6):2269-2273. [doi: 10.1016/j.juro.2012.12.013] [Medline: 23228385]
24. Nwachukwu BU, Adjei J, Trehan SK, Chang B, Amoo-Achampong K, Nguyen JT, et al. Rating a sports medicine surgeon's 'Quality' in the modern era: an analysis of popular physician online rating websites. HSS J 2016 Oct;12(3):272-277 [FREE Full text] [doi: 10.1007/s11420-016-9520-x] [Medline: 27703422]
25. Sobin L, Goyal P. Trends of online ratings of otolaryngologists: what do your patients really think of you? JAMA Otolaryngol Head Neck Surg 2014 Jul;140(7):635-638. [doi: 10.1001/jamaoto.2014.818] [Medline: 24876073]
26. Gilbert K, Hawkins CM, Hughes DR, Patel K, Gogia N, Sekhar A, et al. Physician rating websites: do radiologists have an online presence? J Am Coll Radiol 2015 Aug;12(8):867-871. [doi: 10.1016/j.jacr.2015.03.039] [Medline: 26250980]
27. Riemer C, Doctor M, Dellavalle RP. Analysis of online ratings of dermatologists. JAMA Dermatol 2016 Feb;152(2):218-219. [doi: 10.1001/jamadermatol.2015.4991] [Medline: 26677133]
28. Trehan SK, DeFrancesco C, Nguyen J, Charalel RA, Daluiski A. Online patient ratings of hand surgeons. J Hand Surg Am 2016 Jan;41(1):98-103. [doi: 10.1016/j.jhsa.2015.10.006] [Medline: 26710742]
29. Liu JJ, Matelski JJ, Bell CM. Scope, breadth, and differences in online physician ratings related to geography, specialty, and year: observational retrospective study. J Med Internet Res 2018 Mar 7;20(3):e76 [FREE Full text] [doi: 10.2196/jmir.7475] [Medline: 29514775]
30. Segal J, Sacopulos M, Sheets V, Thurston I, Brooks K, Puccia R. Online doctor reviews: do they track surgeon volume, a proxy for quality of care? J Med Internet Res 2012 Apr 10;14(2):e50 [FREE Full text] [doi: 10.2196/jmir.2005] [Medline: 22491423]
31. Gray BM, Vandergrift JL, Gao G, McCullough JS, Lipner RS. Website ratings of physicians and their quality of care. JAMA Intern Med 2015 Feb;175(2):291-293. [doi: 10.1001/jamainternmed.2014.6291] [Medline: 25437252]
32. Adelhardt T, Emmert M, Sander U, Wambach V, Lindenthal J. Can patients rely on results of physician rating websites when selecting a physician? - A cross-sectional study assessing the association between online ratings and structural and quality of care measures from two German physician rating websites. Value in Health 2015 Nov;18(7):A545. [doi: 10.1016/j.jval.2015.09.1732]
33. Verhoef LM, van de Belt TH, Engelen LJ, Schoonhoven L, Kool RB. Social media and rating sites as tools to understanding quality of care: a scoping review. J Med Internet Res 2014 Feb 20;16(2):e56 [FREE Full text] [doi: 10.2196/jmir.3024] [Medline: 24566844]

34. Murphy GP, Awad MA, Osterberg EC, Gaither TW, Chumnarnsongkhroh T, Washington SL, et al. Web-based physician ratings for California physicians on probation. J Med Internet Res 2017 Aug 22;19(8):e254 [FREE Full text] [doi: 10.2196/jmir.7488] [Medline: 28830852]
35. Daskivich T, Houman J, Fuller G, Black J, Kim H, Spiegel B. Online physician ratings fail to predict actual performance on measures of quality, value, and peer review. J Am Med Inform Assoc 2018 Apr 1;25(4):401-407. [doi: 10.1093/jamia/ocx083] [Medline: 29025145]
36. Alam A, Klemensberg J, Griesman J, Bell C. The characteristics of physicians disciplined by professional colleges in Canada. Open Med 2011;5(4):e166-e172 [FREE Full text] [Medline: 22567070]
37. Alam A, Kurdyak P, Klemensberg J, Griesman J, Bell CM. The characteristics of psychiatrists disciplined by professional colleges in Canada. PLoS One 2012;7(11):e50558 [FREE Full text] [doi: 10.1371/journal.pone.0050558] [Medline: 23209779]
38. Alam A, Khan J, Liu J, Klemensberg J, Griesman J, Bell CM. Characteristics and rates of disciplinary findings amongst anesthesiologists by professional colleges in Canada. Can J Anaesth 2013 Oct;60(10):1013-1019. [doi: 10.1007/s12630-013-0006-8] [Medline: 23897490]
39. Liu J, Alam A, Goldberg H, Matelski J, Bell C. Characteristics of internal medicine physicians disciplined by professional colleges in Canada. Medicine (Baltimore) 2015 Jul;94(26):e937 [FREE Full text] [doi: 10.1097/MD.0000000000000937] [Medline: 26131839]
40. Alam A, Matelski J, Goldberg H, Liu JK, Klemensberg J, Bell CM. The characteristics of international medical graduates who have been disciplined by professional regulatory colleges in Canada: a retrospective cohort study. Acad Med 2017 Feb;92(2):244-249. [doi: 10.1097/ACM.0000000000001356] [Medline: 27603039]
41. Kohatsu ND, Gould D, Ross LK, Fox PJ. Characteristics associated with physician discipline: a case-control study. Arch Intern Med 2004 Mar 22;164(6):653-658. [doi: 10.1001/archinte.164.6.653] [Medline: 15037494]
42. Khaliq AA, Dimassi H, Huang CY, Narine L, Smego RA. Disciplinary action against physicians: who is likely to get disciplined? Am J Med 2005 Jul;118(7):773-777. [doi: 10.1016/j.amjmed.2005.01.051] [Medline: 15989912]
43. RateMDs. URL: https://www.ratemds.com [accessed 2015-11-10]
44. Langevin R, Glancy GD, Curnoe S, Bain J. Physicians who commit sexual offences: are they different from other sex offenders? Can J Psychiatry 1999 Oct;44(8):775-780. [doi: 10.1177/070674379904400803] [Medline: 10566107]
45. George MS. Manitoba workshop provides insight into sexual abuse by physicians. Can Med Assoc J 1993 Mar 1;148(5):815-816 [FREE Full text] [Medline: 8439943]
46. Roback HB, Strassberg D, Iannelli RJ, Finlayson AJ, Blanco M, Neufeld R. Problematic physicians: a comparison of personality profiles by offence type. Can J Psychiatry 2007 May;52(5):315-322. [doi: 10.1177/070674370705200506] [Medline: 17542382]
47. Garfinkel PE, Bagby RM, Waring EM, Dorian B. Boundary violations and personality traits among psychiatrists. Can J Psychiatry 1997 Sep;42(7):758-763. [doi: 10.1177/070674379704200709] [Medline: 9307837]
48. Bismark MM, Spittal MJ, Gurrin LC, Ward M, Studdert DM. Identification of doctors at risk of recurrent complaints: a national study of healthcare complaints in Australia. BMJ Qual Saf 2013 Jul;22(7):532-540 [FREE Full text] [doi: 10.1136/bmjqs-2012-001691] [Medline: 23576774]

Abbreviations

GEEs: generalized estimating equations
IMG: international medical graduate

Edited by T Kool; submitted 16.10.19; peer-reviewed by D Hanauer, T Daskivich, S McLennan; comments to author 10.12.19; revised version received 14.01.20; accepted 12.02.20; published 14.05.20.

Please cite as:
Liu JJ, Goldberg HR, Lentz EJM, Matelski JJ, Alam A, Bell CM
Association Between Web-Based Physician Ratings and Physician Disciplinary Convictions: Retrospective Observational Study
J Med Internet Res 2020;22(5):e16708
URL: https://www.jmir.org/2020/5/e16708
doi:10.2196/16708
PMID:32406851

Original Paper

Disparities in Secure Messaging Uptake Between Patients and Physicians: Longitudinal Analysis of Two National Cross-Sectional Surveys

Dawn M Heisey-Grove[1], MPH, PhD; Henry J Carretta[2], MPH, PhD

[1]Department of Clinical Quality and Informatics, Health Technical Center, MITRE Corporation, McLean, VA, United States
[2]College of Medicine, Florida State University, Tallahassee, FL, United States

Corresponding Author:
Dawn M Heisey-Grove, MPH, PhD
Department of Clinical Quality and Informatics
Health Technical Center
MITRE Corporation
7515 Colshire Dr
McLean, VA, 22102
United States
Phone: 1 7035477389
Email: heiseygrovedm@vcu.edu

Abstract

Background: Emails securely exchanged between patients and clinicians offer the promise of improved access to care and indirectly improved health outcomes. Yet research to date is mixed on who—among both patients and clinicians—is using secure messaging.

Objective: Using data from two large nationally representative cross-sectional surveys, this study aimed to compare the prevalence of secure messaging use among patients and their access to the functionality through their physicians, and to explore the clinical practice and physician characteristics and patient sociodemographic characteristics associated with the use of secure messaging.

Methods: We conducted regression analyses to identity statistical associations between self-reported secure messaging use and access, and the patient, practice, and physician characteristics from the National Health Interview Survey (NHIS) and the National Ambulatory Medical Care Survey (NAMCS). The NHIS data collected between 2013 and 2018, with approximately 150,000 adult individuals, were used to evaluate patient characteristics associated with email communication with clinicians. The NAMCS data included 7340 physicians who reported on secure messaging use between 2013 and 2016 and provided context on physician specialty, use of certified health information technology (IT), and practice size and ownership associated with secure messaging access and use.

Results: By 2016, two-thirds of ambulatory care visits were conducted by a physician who reported using secure messaging, up from 40.70% in 2013. The percentage of US residents who reported sending an email to their clinician, however, only increased from 7.22% to 16.67% between 2013 and 2018. We observed a strong positive association between certified health IT use and secure messaging use (odds ratio [OR] 11.46, 95% CI 7.55-17.39). Individuals who were black, had lower levels of education, had Medicaid or other public payer insurance, or those who were uninsured had reduced odds for using email to communicate with clinicians. No differences were observed in secure messaging use based on physician specialty, but significant differences were observed by practice size (OR 0.46, 95% CI 0.35-0.60 in solo practices vs nonsolo practices) and practice ownership (P<.001 for the different categories).

Conclusions: This study is the first to use two large nationally representative surveys to produce longitudinal estimates on the access and use of patient-clinician email communication in the United States. The survey findings complement each other: one provides the patient perspective of their use and the other indicates potential patient access to secure messaging based on the use of the functionality by the physicians providing treatment. This study provides nationally representative data on the characteristics of patients and physicians who have access to and are using secure messaging. This information can be used to target interventions to promote adoption and use of secure messaging.

(J Med Internet Res 2020;22(5):e12611) doi:10.2196/12611

KEYWORDS

electronic mail; health communication; electronic health records

Introduction

Conceptually, expanding patient and clinician communications beyond health care facility walls improves patients' access to care by providing a forum for patients to get answers to their questions without requiring in-person visits [1,2]. One mechanism to expand communication is through forms of computer-mediated communication such as email. Although early mechanisms to exchange email between patient and clinician were less secure, it is now common for email exchange to be conducted using a secure patient portal. Dubbed "secure messaging," this form of communication is defined by the Centers for Medicare & Medicaid Services as "any electronic communication between a provider and patient that ensures only those parties can access the communication. This electronic message could be email or the electronic messaging function of a personal health record, an online patient portal, or any other electronic means" [2].

In 2001, the Institute of Medicine (IOM) noted that patient-provider communication via email had the potential to reduce costs while meeting patient needs more quickly [1]. Both patients and clinicians identified benefits of secure messaging, which included convenience, not feeling rushed, improved patient access, more direct and focused communication, increased efficiency, avoidance of phone tag, improved communication between visits, and improved patient engagement, satisfaction, and trust [3-7].

Since the publication of the IOM report, most physician practices have adopted secure messaging functionality, although these studies measured the number of physicians rather than the proportion of patients who have access to the functionality via their physicians [8-11]. Patients expressed interest in sending messages to their clinicians if given the opportunity, and they were receptive to receiving and reading the messages sent to them: the vast majority of messages sent to patients were read within 3 days, and fewer than 5% were not read within 3 weeks [12]. Yet a study by Tarver et al [13] estimated that only 3 in 10 individuals reported communicating with their clinicians using email or the internet in 2013.

In 2014, the Medicare and Medicaid electronic health record (EHR) Incentive Programs required that eligible professionals participating in stage 2 of the program use secure messaging to communicate with their patient population (not just Medicare beneficiaries) [2]. To receive incentive payments and avoid penalties, participating professionals had to use EHR systems that met meaningful use criteria defined by the Department of Health and Human Services (ie, certified health information technology [IT]). We would therefore expect that availability of the EHR functionality supporting secure messaging would increase as the 2014 requirement approached and continue to increase as more providers met the stage 2 criteria, and with similar requirements that were included in the first 2 years of the subsequent program (the Medicare Access and CHIP Reauthorization Act of 2015 [MACRA]).

Clinicians' patterns of secure message communication affect patients' use of the functionality: patients were more likely to initiate messages if their clinicians responded quickly and had a higher overall response rate [14]. Patients whose clinicians initiated more message threads were also more likely to initiate their own threads. We would therefore expect that patients' use of secure messaging would increase as their access to that functionality via their providers' infrastructure increased.

We used data from two nationally representative surveys to explore whether patients' use of email exchange paralleled the availability of secure messaging functionality in ambulatory care settings. We present data pertaining to patients' self-reported use of email to communicate with their clinicians and patients' potential access to that functionality through physician visits.

Methods

Overview

Data from two nationally representative surveys were analyzed. In this section, the methodology associated with each survey is reported separately, starting first with methodologic approaches using data from the National Ambulatory Medical Care Survey (NAMCS), with which the proportion of office visits with access to secure messaging functionality were estimated. Approaches to using data from the National Health Interview Survey (NHIS) helped estimate the prevalence of patients who exchanged email with clinicians.

National Ambulatory Medical Care Survey

The NAMCS is conducted annually by the National Center for Health Statistics (NCHS) using a nationally representative sample of ambulatory care medical visits to nonfederally employed office-based physicians. The NAMCS is an annual cross-sectional survey designed to provide estimates of ambulatory care medical visits to office-based physicians in the United States [15-19]. The survey captures information at both the individual physician and patient visit levels. Beginning in 2012, the NAMCS asked physicians if they regularly used, or had the functionality to send secure electronic messages (eg, email) to their patients, and in 2013, physicians were asked if they used certified health IT. We therefore analyzed NAMCS data from 2013 through 2016 to develop yearly snapshots of the availability of secure messaging to physicians and patients in the context of the ambulatory care medical visits.

National Ambulatory Medical Care Survey Study Population

The unit of measurement for the NAMCS was the patient-physician visit, excluding telephone consults, hospital visits, house calls, institutional settings, other visits performed outside the physician's office, and visits made solely for administrative purposes (eg, leaving a specimen, paying a bill) [15,20].

NAMCS data were collected from physicians included in master lists from the American Medical Association and American Osteopathic Association who met the following criteria: office-based as defined by the respective association, principally engaged in patient care activities, and less than 85 years at the time of survey [15-19]. Physicians who were primarily employed in federal institutions or who had specialties of anesthesiology, pathology, or radiology were excluded. Each physician who met the criteria was assigned a random week during the year; if the physician saw no patients during that time period (eg, due to vacation or illness), he or she was excluded from the sample. In addition, the physician was excluded if patient visit data for the assigned time period was not recorded. Therefore, NAMCS data should not be considered representative of all ambulatory care office-based physicians. In our analyses, we used the prevalence of visits based on the physicians' self-reported characteristics because this leverages the survey data as designed; it also allows us to better approximate the availability of secure messaging to patients with visits to physicians who can offer that functionality to communicate between visits.

Physician sampling was stratified based on census region, state, doctor type (Doctor of Medicine or Doctor of Osteopathy), practice type, metropolitan statistical area, and 14 specialty categories (Multimedia Appendix 1) [15]. Table 1 displays the number of respondents and unweighted response rates for the years included in our analyses.

Table 1. National Ambulatory Medical Care Survey response rates and number of sampled office visits, 2013-2016.

Survey year	Eligible physician respondents (unweighted response rate[a]), n (%)	Ambulatory visits, n
2016	721 (32.37)	13,165
2015	1415 (28.82)	28,332
2014	2325 (38.64)	45,710
2013	2879 (41.13)	54,873

[a]Conditional response rates reported for families, sample adults, and children; total response rate reported for households. Data in tables based on author's compilation of information from data from the appropriate survey description documents published by the National Center for Health Statistics [15-17,19,21].

Dependent Variable From the National Ambulatory Medical Care Survey

Our dependent variable in NAMCS-related analyses was whether the physician conducting the ambulatory visit reported access to secure messaging. The question in the NAMCS was not specific to the visit itself; rather, it was related to the physician's access to and use of (for surveys in 2013 through 2015) secure messaging functionality. This survey question changed with the 2016 survey to a dichotomous variable related to access to the secure messaging functionality, which is how we present the data in our analyses.

Independent Variables From the National Ambulatory Medical Care Survey

Our NAMCS-based analyses included both physician and practice characteristics. We included only one physician-level characteristic: physician specialty was identified based on data provided to the NCHS by the American Medical Association and the American Osteopathic Association [15]. We opted to use the NAMCS-defined categories for clinical specialty: primary care, medical specialty, and surgical specialty.

Practice characteristics included the use of certified health IT and practice size, ownership, and geographic region. We measured the use of certified health IT as an affirmative response to whether the physician's current information system met the meaningful use criteria defined by the Department of Health and Human Services, which refers to the health IT certification criteria that support the Medicare and Medicaid Electronic Health Record Incentive programs (eg, *meaningful use*). Physicians who did not have an electronic system were coded as having no access to secure messaging.

Practice size and ownership were based on the practice in which each visit was conducted. We used a dichotomous practice size variable (solo/nonsolo). We included three categories for practice ownership: physician or physician group; insurance company or health plan; and academic medical center, community health center, or other hospital.

The geographic region was based on the location of the office in which the visit occurred and was categorized into census regions: Midwest, Northeast, South, and West. For all variables, nonresponse and unknown values were coded as missing.

National Health Interview Survey

The NHIS, which was the second survey we analyzed for this research, is conducted annually by the NCHS using a nationally representative sample of civilian US residents. Beginning in 2011, the NHIS included a question asking adults if they communicated with their clinicians via email. Pairing responses to this question with sociodemographic information gathered through the NHIS provides a more comprehensive view of those in the United States who reported communicating with their clinicians using email. We included survey data only from 2013 through 2018 because the *internet use* variable was added in 2013, and internet access is a critical factor in patients' ability to access secure messaging.

Similar to the NAMCS, the NHIS is an annual cross-sectional survey that uses a nationally representative sample of US residents selected based upon a complex, multistage-stratified sampling process [22,23]. The NHIS consists of four core components that capture information on the household, families within the sampled household, and a randomly selected child (when available) and adult from each family. Questions in the survey's core components were asked consistently across survey

years; Multimedia Appendix 1 lists the survey questions used for this study.

National Health Interview Survey Study Population

NHIS study participants were randomly selected to represent the US population with oversampling for blacks, Asians, and adults older than 65 years within the sampled adults [23]. Excluded from the NHIS were individuals in long-term care and correctional facilities, as well as individuals living outside the country. Households with members of the Armed Forces were included only if at least one member in the family was not in the Armed Forces, in which case results only from the household members not in the Armed Forces were used in the final analyses.

Sampling was based on households, with approximately 36,000 households targeted each year to achieve a survey goal of about 87,500 individuals each year. In-person interviews were conducted with adults (individuals older than 17 years) in each household; this representative provided information about all members of the household. Each household was further subdivided into families with a responsible adult respondent for each family, and sampled children and adults (one of each per family). For each family sampled by NHIS, a sample adult was randomly selected to complete the *sample adult* questionnaire. If that sampled adult was absent at the time of the interview, the responsible adult family member could provide answers for them. We based our analyses on these sampled adults. Table 2 displays the survey response rate for each of the survey years we included in our analyses.

Table 2. Number of households, families, and adults included in each National Health Interview Survey, 2013-2017.

Respondent type	Survey (response rate[a])					
	2018, n (%)	2017, n (%)	2016, n (%)	2015, n (%)	2014, n (%)	2013, n (%)
Households	29,839 (64.16)	32,617 (66.47)	40,220 (67.90)	41,493 (70.12)	44,552 (73.83)	41,335 (75.69)
Families	30,309 (98.73)	33,157 (98.90)	40,875 (98.85)	42,288 (98.89)	45,497 (98.98)	42,321 (98.96)
Sample adults	25,417 (83.89)	26,742 (80.69)	33,028 (80.86)	33,672 (79.66)	36,697 (80.54)	34,557 (81.71)

[a]Conditional response rates reported for families, sample adults, and children; total response rate reported for households. Data in tables based on author's compilation of information from data from the appropriate survey description documents published by the National Center for Health Statistics [22,24-30].

Dependent Variable From National Health Interview Survey

Our dependent variable for NHIS-based analyses was individuals' self-reported email communication with clinicians. The NHIS captures this element with five options: Yes, No, Refused, Not ascertained, and Don't know. We created a dichotomous variable (yes, no) to capture this response and excluded individuals from analyses if their response was Refused, Not ascertained, or Don't know.

Independent Variables From National Health Interview Survey

For NHIS analyses, we included age as a categorical variable (18-44 years, 45-64 years, and 65 years and older), sex, race (white, African American/black, and other race), Hispanic (yes, no), and census region (Northeast, Midwest, South, and West). On the basis of the literature about patient factors relevant to patient-centered communication that promotes improved outcomes [31,32], we included the variables for education and familiarity in speaking the English language in our analyses. Our education variable included six categories: less than a high school education, some college education but no degree, associate's degree, bachelor's degree, and graduate or professional degree. To assess individuals' comfort with the English language, which is particularly relevant for written communication and patients' comfort level when communicating with clinic staff [33,34], we used the responses to the NHIS question How well is English spoken, which included four categories (very well, well, not well, and not at all).

We also included a variable as a proxy for health care access. Our health insurance variable included five categories for private insurance, Medicare, Medicaid or other public insurance, military insurance, and uninsured. In addition, because we presented these data against those from a physician-based survey, we included a variable that assessed whether the individual saw or spoke with clinic staff in the 12 months preceding the survey. We created this as a categorical variable that distinguished between contact with a physician and contact with other clinician types. We created the physician's designation by consolidating three NHIS questions about whether the patient saw a doctor who specializes in women's health, a medical doctor who specializes in a particular medical disease or problem, and a general doctor who treats a variety of illnesses. The category for other clinician type included nurse practitioners, midwives, physician assistants, therapists, chiropractors, podiatrists, optometrists, ophthalmologists, or mental health professionals.

The ability to communicate with clinic staff using email depends on individuals' access to and use of the internet. We therefore included a dichotomous variable to assess individuals' use of the internet. Finally, we included a categorical variable to account for the survey year, one for each of the 6 years included in the analyses (2013 through 2018).

For all variables, unknown and nonresponses were coded as missing and those individuals excluded from the analyses.

Statistical Analyses for Both Surveys

Our first regression model used NAMCS data, including survey year (2013 through 2016), practice characteristics (solo practice, practice ownership, and use of certified health IT), and physician

specialty. The dependent variable for these analyses was the physician's reported use of secure messaging. Our second regression model—based on NHIS data—included individuals' characteristics (such as age, sex, race, ethnicity, geographic region, education, English language, and health insurance type), internet use, whether they saw or spoke with a clinician in the year preceding the survey, and survey year (2013 through 2018). The dependent variable for the NHIS analyses was self-reported email communication with physicians.

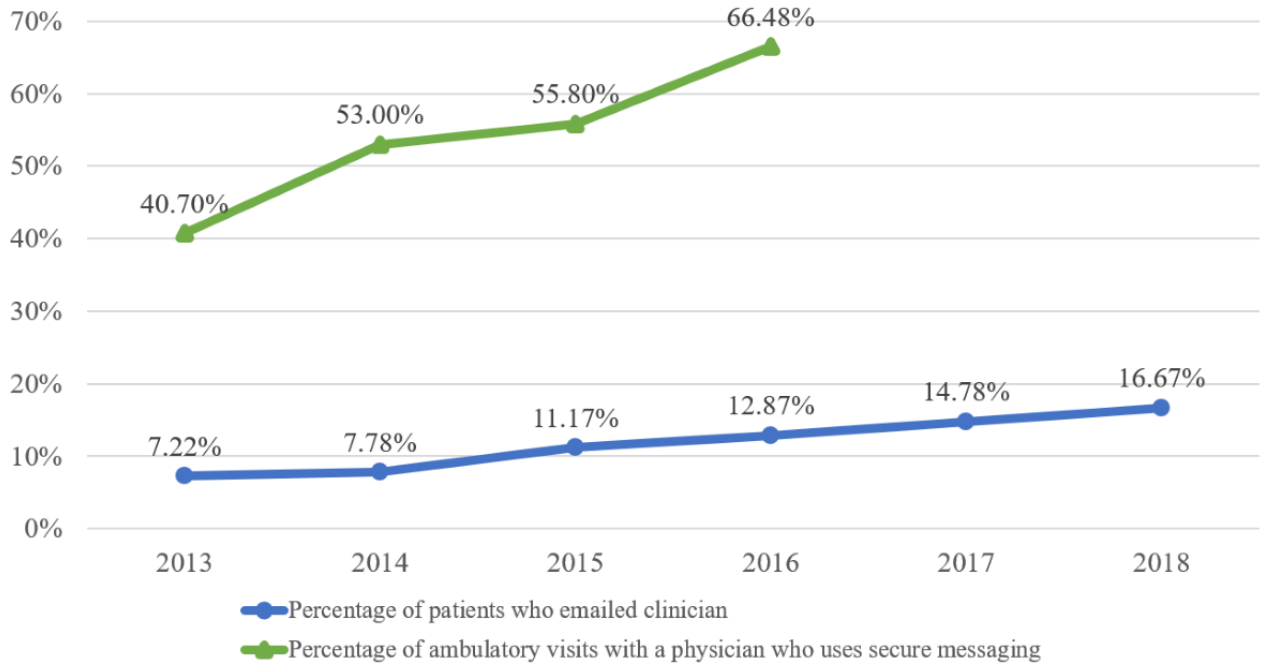
All analyses accounted for the complex sampling techniques used by NCHS by leveraging sample weights for stratification and primary sampling units. We estimated unadjusted statistical differences by year and individuals' or physicians' characteristics using chi-square test. We performed logistic regression to estimate associations between characteristics and the use of email or secure messaging. We used casewise deletion for all missing values (see Multimedia Appendix 2 for tables of missing values for each survey). Analyses were conducted using SAS software version 9.4 (SAS Institute Inc, Cary, NC).

Results

Email Availability and Use Across Both Surveys

Figure 1 displays prevalence data of email use by patients and physicians from both NHIS and NAMCS. Between 2013 and 2016, the percentage of ambulatory care visits with a physician who reported using secure messaging with patients increased 63% for a high of 66.48% in 2016. Although there was a larger percent increase (132%) between 2013 and 2018 of patients reporting the use of email to communicate with their clinicians, by 2018 only 16.67% of US residents reported using email to communicate with their clinicians.

Figure 1. Prevalence of email use and access among US residents and ambulatory care visits, 2013-2018 (on the basis of authors' analysis of National Center for Health Statistics [NCHS] and National Health Interview Surveys [NHIS], 2013-2018 [patients who emailed their clinician] and NCHS National Ambulatory Medical Care Surveys, 2013-2016 [ambulatory visits with physicians who used secure messaging]). Percentages are weighted national estimates.



Physician and Practice Characteristics (National Ambulatory Medical Care Survey)

Figure 2 provides the percentages of ambulatory visits by physician specialty and use of secure messaging in 2016. A total of 70.19% of ambulatory medical care visits were conducted by primary care physicians who used secure messaging (38.28% of all visits). Two-thirds of surgical specialist visits were conducted by surgical specialists who used secure messaging (12.69% of all visits), while fewer than six in 10 medical specialist visits (15.51% of all visits) were conducted by those specialists who used secure messaging.

Figure 3 displays the proportion of ambulatory care visits stratified by physicians' use of certified health IT products and secure messaging use. A total of 83.62% of physician visits were conducted using certified health IT products, and three-quarters of those (or 63.04% of all visits) were conducted by physicians who reported secure messaging use. Of the 16.38% of ambulatory medical visits conducted by physicians not using certified health IT, only about 2 in 10 (or 3.32% of all visits) were performed by physicians using secure messaging.

Figure 2. Percentage of ambulatory care visits by physicians' specialty and secure messaging use, 2016 (on the basis of authors' analysis of National Center for Health Statistics and National Ambulatory Medical Care Survey, 2015). Percentages are weighted national estimates of ambulatory care visits.

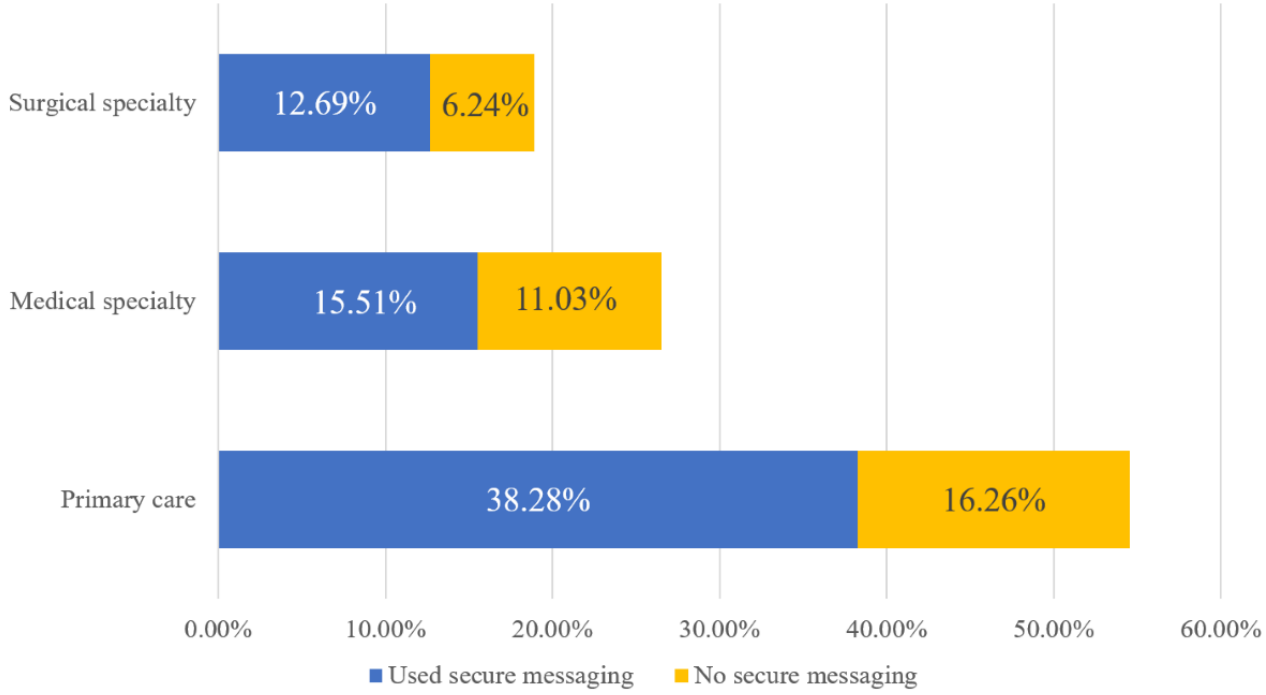


Figure 3. Percentage of ambulatory care visits by physicians' use of certified health IT products and secure messaging use, 2016 (on the basis of authors' analysis of National Center for Health Statistics and National Ambulatory Medical Care Survey, 2016). Percentages are weighted national estimates of ambulatory care visits.

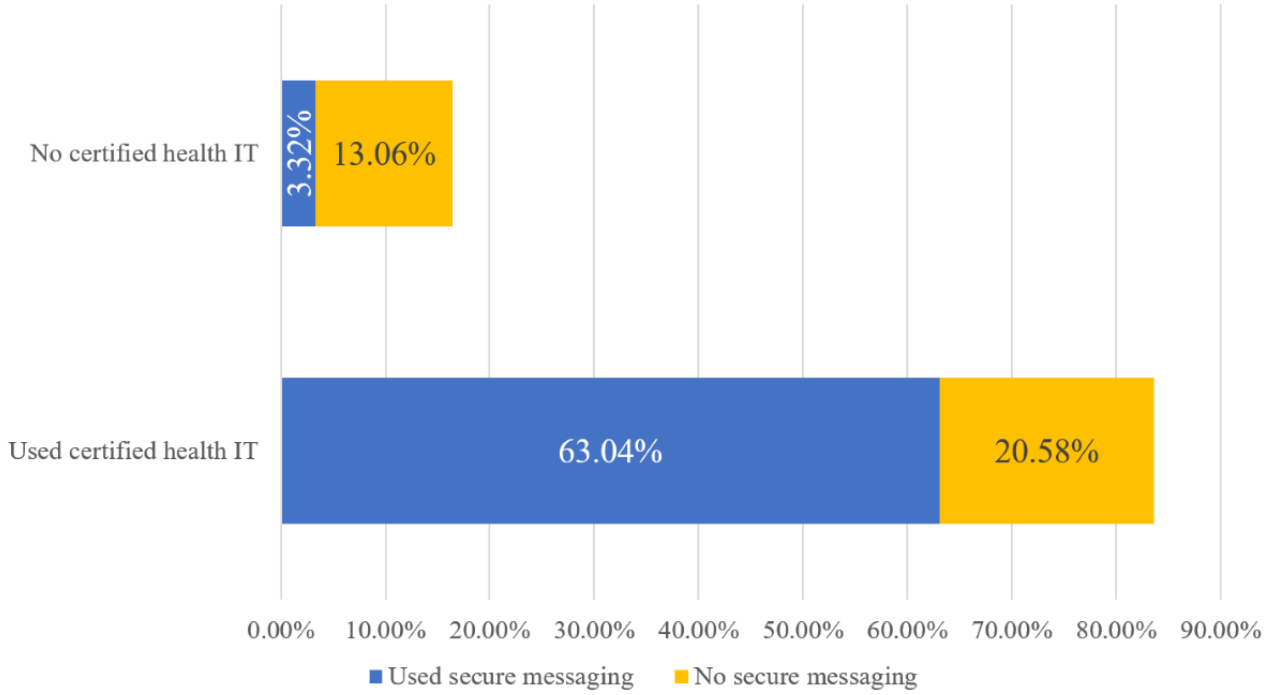


Figure 4 displays the percentages of ambulatory medical care visits conducted by physicians using secure messaging based on practice ownership. More than 90% of all ambulatory visits conducted in health maintenance organization (HMO) or insurance-owned practices and academic medical centers or hospital-owned organizations were conducted by physicians who used secure messaging. In contrast, slightly less than two-thirds of all visits at physician-owned practices were conducted by physicians who used secure messaging (47.88% of all ambulatory medical care visits).

Figure 4. Percentage of ambulatory care visits by practice ownership and secure messaging use, 2016 (on the basis of authors' analysis of National Center for Health Statistics and National Ambulatory Medical Care Survey, 2016). Percentages are weighted national estimates of ambulatory care visits. CHC: community health center; HMO: Health Maintenance Organization.

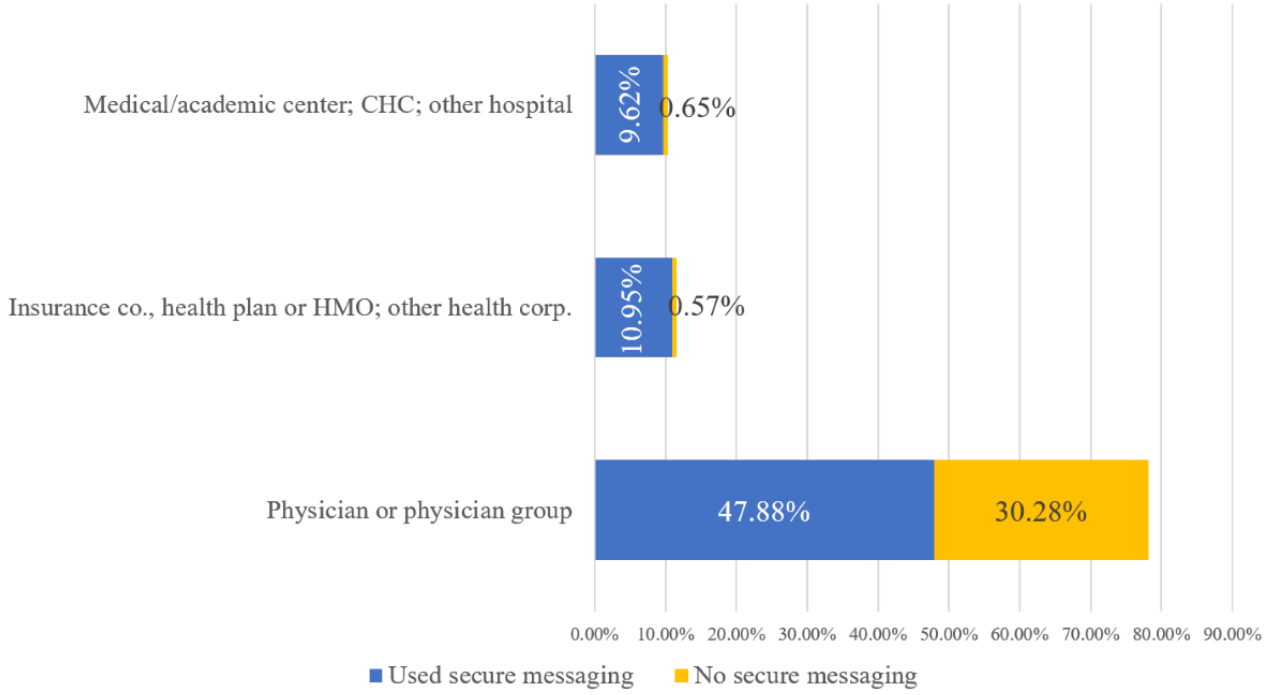
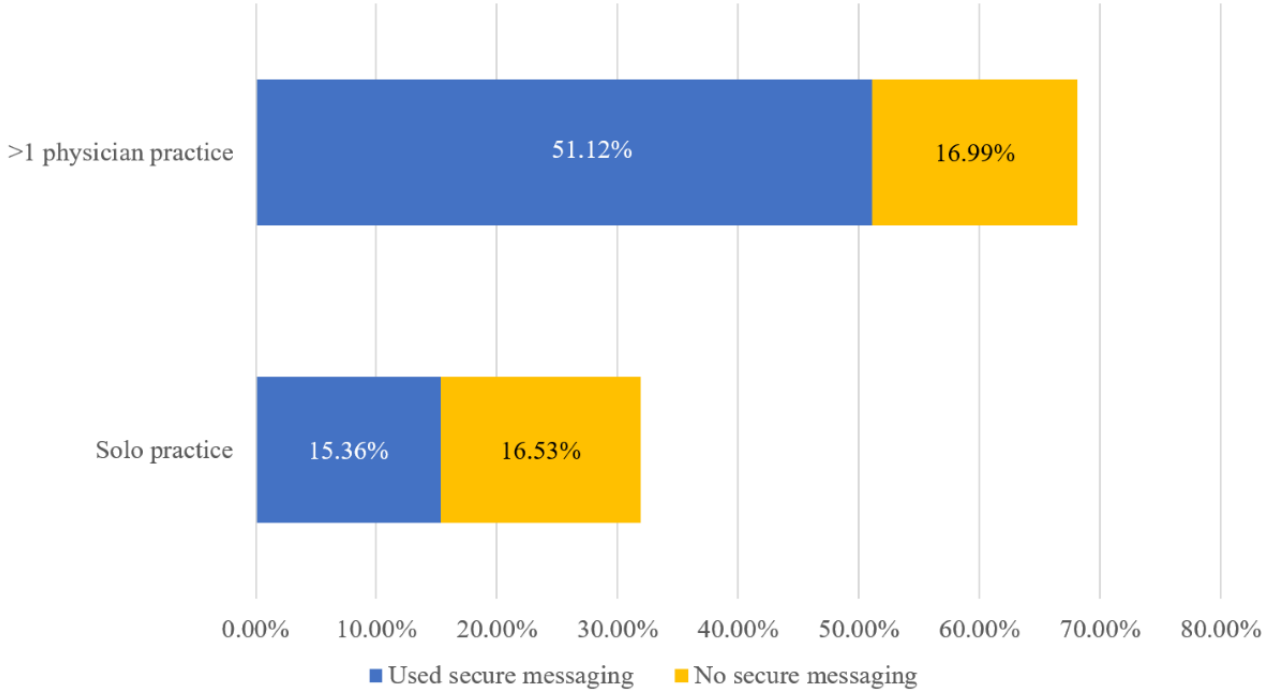


Figure 5 displays the percentage of secure messaging use based on practice size. Slightly less than half of ambulatory care visits in solo physician practices (15.36% of all visits) were conducted by physicians who had access to secure messaging functionality. In contrast, three-quarters of ambulatory visits in practices with more than 1 physician were conducted by physicians who used secure messaging (51.12% of all visits).

Figure 5. Percentage of ambulatory care visits by practice size and physicians' reported secure messaging use, 2016 (on the basis of authors' analysis of National Center for Health Statistics and National Ambulatory Medical Care Survey, 2016). Percentages are weighted national estimates of ambulatory care visits.).



Multimedia Appendix 3 presents a comparison between 2013 and 2016 of the prevalence of secure messaging use by physician and practice characteristics. Across all characteristics, there was a statistically significant increase over the 4 years, which persisted in adjusted analyses. Table 3 presents the adjusted estimates of association, controlling for physician and practice

characteristics. No differences were observed by specialty. Physicians using certified health IT had greater odds of using secure messaging (OR 11.46, 95% CI 7.55-17.39) than physicians not using certified products. Ambulatory care visits in solo physician practices had lower odds of being conducted by physicians who used secure messaging (OR 0.46, 95% CI 0.35-0.60) than those in practices with more than 1 physician. Visits conducted in physician-owned practices had lower odds of being conducted by physicians using secure messaging than practices owned by other entities.

Table 3. Regression results for association between ambulatory care with secure messaging functionality and physician and practice characteristics, based on authors' analysis of National Center for Health Statistics, National Ambulatory Medical Care Surveys, 2013-2016.

Characteristic and independent variables	OR[a] (95% CI)	*P* value
Certified health IT[b] use		
Yes	11.46 (7.55-17.39)	<.001
No	Referent	Referent
Solo physician practice		
Yes	0.46 (0.35-0.60)	<.001
No	Referent	Referent
Practice ownership		
Insurance company, health plan, or HMO[c]; other health corporation	1.81 (1.29-2.54)	<.001
Medical/academic health center; CHC[d]; other hospital	2.12 (1.57-2.87)	<.001
Physician or physician group	Referent	Referent
Physician specialty		
Medical	0.83 (0.62-1.11)	.20
Surgical	0.90 (0.70-1.16)	.42
Primary care	Referent	Referent
Region		
Midwest	0.85 (0.61-1.18)	.32
Northeast	0.55 (0.38-0.80)	<.01
South	0.94 (0.66-1.34)	.74
West	Referent	Referent
Survey year		
2013	0.29 (0.21-0.42)	<.001
2014	0.55 (0.38-0.79)	<.01
2015	0.60 (0.40-0.90)	.01
2016	Referent	Referent

[a]OR: odds ratio.

[b]IT: information technology.

[c]HMO: health maintenance organization.

[d]CHC: community health center.

Patient Characteristics (National Health Interview Survey)

Table 4 lists the 2013 and 2018 reported rates of email communication with clinicians by individuals' sociodemographic characteristics. Prevalence of email communication increased across all categories except among individuals who did not speak English well or at all. We observed large percent increases across the 6 years among individuals with Medicare (236% change from 4.05% to 13.63%) and military insurance (154% change from 9.06% to 23.03%), and those who were uninsured (146% change from 1.91% to 4.71%). We also found a large increase in communication via email among Hispanic individuals (180% change from 3.36% to 9.41%). By education level, the largest percent change was among individuals with a high school diploma or its equivalent (165% change from 3.26% to 8.66%).

Table 4. Prevalence of email communication with clinicians by individuals' characteristics based on authors' analysis of National Center for Health Statistics, National Health Interview Surveys, 2013 and 2018. Percentages are weighted national estimates.

Characteristics	Individuals emailing clinicians in 2013, % (95% CI)	Individuals emailing clinicians in 2018, % (95% CI)	*P* value
Age (years)			**<.001**
18-44	6.82 (6.25-7.39)	15.98 (14.87-17.10)	
45-64	8.81 (8.12-9.50)	19.15 (17.97-20.33)	
65+	5.23 (4.47-6.00)	14.21 (13.09-15.32)	
Sex			**<.001**
Male	6.02 (5.47-6.57)	14.25 (13.29-15.22)	
Female	8.33 (7.77-8.89)	18.93 (17.83-20.02)	
Education			**<.001**
Less than a high school diploma	1.00 (0.65-1.36)	2.55 (1.84-3.26)	
High school diploma or equivalent	3.26 (2.74-3.79)	8.66 (7.74-9.58)	
Some college, no degree	6.54 (5.69-7.39)	15.21 (13.74-16.68)	
Associate's degree	6.97 (5.92-8.01)	17.41 (15.72-19.09)	
Bachelor's degree	12.87 (11.87-13.86)	25.18 (23.51-26.86)	
Graduate or professional degree	17.06 (15.49-18.64)	33.23 (31.03-35.43)	
Hispanic			**<.001**
Yes	3.36 (2.78-3.93)	9.41 (8.00-10.83)	
No	7.91 (7.50-8.31)	18.09 (17.23-18.95)	
Race			**<.001**
Black	4.68 (3.85-5.50)	11.07 (9.74-12.40)	
Other race	9.07 (7.71-10.43)	16.86 (14.17-19.56)	
White	7.45 (7.04-7.86)	17.56 (16.64-18.48)	
Speak English			
Not at all	0.30 (0.00-0.77)	0.48 (0.00-1.15)	.67
Not well	1.13 (0.58-1.68)	0.75 (0.16-1.34)	.35
Well	3.51 (2.49-4.54)	9.48 (7.40-11.55)	<.001
Very well	8.39 (7.75-9.03)	18.16 (17.29-19.04)	<.001
Region			**<.001**
Northeast	6.35 (5.57-7.14)	15.59 (13.77-17.42)	
Midwest	6.05 (5.25-6.84)	16.28 (14.78-17.78)	
South	5.88 (5.37-6.40)	14.54 (13.28-15.80)	
West	11.17 (10.29-12.06)	21.10 (18.78-23.42)	
Health insurance			**<.001**
Uninsured	1.91 (1.48-2.34)	4.71 (3.70-5.73)	
Medicaid or other public payer	3.76 (2.90-4.62)	7.70 (6.48-8.92)	
Medicare	4.05 (3.32-4.78)	13.63 (12.14-15.12)	
Military	9.06 (6.92-11.19)	23.03 (18.35-27.72)	
Private	9.61 (9.08-10.13)	20.54 (19.50-21.57)	
Saw/spoke to clinician in prior 12 months			**<.001**

Characteristics	Individuals emailing clinicians in 2013, % (95% CI)	Individuals emailing clinicians in 2018, % (95% CI)	*P* value
Physician	8.81 (8.36-9.27)	20.01 (19.03-20.99)	
Nonphysician clinician	4.85 (3.65-6.04)	9.23 (7.53-10.93)	
None	1.54 (1.1-1.98)	3.34 (2.54-4.14)	
Internet use			**<.001**
Yes	9.38 (8.93-9.84)	20.09 (19.13-21.04)	
No	0.44 (0.23-0.65)	1.28 (0.85-1.71)	

In 2018, the highest percentage of individuals using email to communicate with their clinicians lived in the West (21.10%), saw or spoke with their physician in the preceding 12 months (20.01%), and were between the ages of 45 and 64 years (19.96%). Almost twice as many non-Hispanics reported email communication compared with Hispanics (18.09% vs 9.41%). Fewer black individuals (11.07%) reported using email to communicate with their clinicians compared with white individuals (17.56%).

Table 5 lists multivariate analysis results using NHIS data from 2013 through 2018. Individuals with a graduate or professional degree had greater odds of communicating with clinicians using email than individuals with less education. The English language was also positively associated with email communication, with individuals who spoke English very well having greater odds for email communication with clinicians than individuals who English language skills were less well developed. After controlling for other variables in the model, we found no statistical difference between the youngest individuals and oldest individuals (*P*=.57), nor between white individuals and those of other races (*P*=.73). Black race was associated with reduced odds of communicating using email (OR 0.83, 95% CI 0.77-0.90). Hispanic individuals had lower odds for communicating with clinicians using email than non-Hispanic individuals (OR 0.79, 95% CI 0.73-0.86).

We observed large odds for email use among individuals who saw or spoke with a physician (OR 4.81, 95% CI 4.31-5.36) or nonphysician clinician (OR 2.19, 95% CI 1.89-2.53) in the preceding year. Uninsured individuals, and those with Medicaid or other public payers, had lower odds of communicating via email with clinicians than individuals with private payers; however, we found no statistical difference between individuals with private payers and individuals with either military insurance or Medicare.

Differences by region mirrored what we observed in the regional differences from the NAMCS analyses. As expected, internet use was strongly associated with email communication with clinicians (OR 11.10, 95% CI 9.39-13.11).

Table 5. Adjusted associations between individuals' characteristics and email communication with clinicians based on authors' analysis of National Center for Health Statistics, National Health Interview Surveys, 2013-2018.

Independent variable	OR[a] (95% CI)	*P* value
Age (years)		
18-44	1.02 (0.95-1.09)	.57
45-64	1.24 (1.16-1.32)	<.001
65 years and older	Referent	Referent
Sex		
Female	0.80 (0.77-0.84)	<.001
Male	Referent	Referent
Education		
Less than a high school diploma	0.18 (0.16-0.21)	<.001
High school diploma or equivalent	0.29 (0.27-0.31)	<.001
Some college, no degree	0.43 (0.40-0.47)	<.001
Associate's degree	0.46 (0.43-0.50)	<.001
Bachelor's degree	0.71 (0.67-0.76)	<.001
Graduate or professional degree	Referent	Referent
Hispanic		
Yes	0.79 (0.73-0.86)	<.001
No	Referent	Referent
Race		
Black	0.83 (0.77-0.90)	<.001
Other race	0.98 (0.90-1.08)	.73
White	Referent	Referent
Speak English		
Not at all	0.29 (0.17-0.51)	<.001
Not well	0.30 (0.22-0.41)	<.001
Well	0.73 (0.64-0.84)	<.001
Very well	Referent	Referent
Region		
Northeast	0.50 (0.45-0.57)	<.001
Midwest	0.62 (0.55-0.68)	<.001
South	0.58 (0.52-0.64)	<.001
West	Referent	Referent
Health insurance		
Uninsured	0.52 (0.46-0.58)	<.001
Medicaid or other public payer	0.65 (0.59-0.71)	<.001
Medicare	1.01 (0.93-1.09)	.88
Military	1.11 (0.97-1.28)	.13
Private	Referent	Referent
Saw/spoke to clinician in prior 12 months		
Physician	4.81 (4.31-5.36)	<.001
Nonphysician clinician	2.19 (1.89-2.53)	<.001
None	Referent	Referent

Independent variable	OR[a] (95% CI)	*P* value
Internet use		
Yes	11.10 (9.39-13.11)	<.001
No	Referent	Referent
Survey year		
2013	0.43 (0.39-0.48)	<.001
2014	0.42 (0.39-0.46)	<.001
2015	0.63 (0.59-0.69)	<.001
2016	0.74 (0.69-0.79)	<.001
2017	0.86 (0.80-0.92)	<.001
2018	Referent	Referent

[a]OR: odds ratio.

Discussion

Principal Findings

This research leverages data from two large nationally representative surveys to provide longitudinal prevalence rates for the use of, and access to, email to communicate with clinicians. Between 2013 and 2016, we observed a statistically significant increase in ambulatory medical care visits conducted by physicians who used secure messaging, such that by 2016, two-thirds of all visits were administered by physicians who used secure messaging. We observed significant positive associations between certified health IT use and secure messaging use.

The strong association of secure messaging use with certified health IT may be attributed to the requirements of the meaningful use program, and its subsequent replacement program (MACRA), for eligible providers to report on secure messaging use. Because it takes time to implement workflows to support IT changes, the increase between 2013 and 2014 demonstrated in this study may be associated with physicians increasing their access and use of secure messaging in anticipation of the stage 2 requirements. Moreover, visits conducted by physicians without certified health IT had lower odds of having secure messaging functionality. Although our data do not provide a cause-and-effect answer, the timing and degree of the increases give credence to the idea that much of the uptake by physicians was due to meaningful use and MACRA. Because secure messaging has shown to be positively associated with patient health outcomes [35-42], driving physicians' behavior through similar programs and regulations may be important for policy makers to consider as they seek ways to increase patient engagement and promote improved health outcomes.

In contrast to the two-thirds ambulatory medical care visits with secure messaging access in 2016, 2 years later in 2018, only 17% of US residents reported communicating with their clinicians using email. Individuals who saw or spoke with a physician in the preceding year had greater odds of communicating via email with a clinician during that time period. Clinicians' secure messaging use has been demonstrated to influence patients' messaging behaviors [14,43]. Our findings demonstrate that most physicians have secure messaging capabilities, but patients were not taking advantage to communicate using that modality with this clinic staff. It seems that access to secure messaging functionality alone may be insufficient to change behaviors.

Many patients expressed intention to send messages to their clinicians if given the opportunity [5,33]. In addition, patients seemed receptive to receiving and reading the messages sent to them: the vast majority of messages sent to patients were read within 3 days, and fewer than 5% of messages were not read within 3 weeks [12]. Therefore, there may be a need for physicians to encourage patients to use the secure messaging forum to communicate to if we wish to see an increase in patients' use of secure messaging.

There is moderate supporting evidence of associations between message use and selected patient outcomes (eg, glucose levels in patients with diabetes), and some evidence for other outcomes (eg, diastolic and systolic blood pressure among patients with hypertension) [44]. We observed significant differences in email communication by individuals based on education, race, ethnicity, and insurance status, with patients with lower levels of education, black patients, those with Medicaid or other public payers, and uninsured patients having reduced odds for secure messaging use. Such differences in the use of a communication modality that might have positive impacts on health outcomes—which permits patients to communicate with clinic staff at their convenience and can increase satisfaction and improve understanding of their condition [6,45]—may further exacerbate health disparities if not addressed.

The discrepancies in the proportion of ambulatory care visits associated with access to and use of secure messaging use by practice size and ownership may be due to limited staffing resources of solo and physician-owned practices. Clinical responses to patient-generated messages were frequently triaged through a clinical response team that might include nurses (registered, licensed practical, or advanced practice), physician assistants, pharmacists, and physicians [46-48]. Effective workflow design may be critical to gaining acceptance of secure messaging among clinical teams because workflows facilitating

this team-based approach to response may be complicated and confusing [48]. Development of strategies and resources that are effective for less well-resourced physician practices may be critical to promoting secure messaging use among patients.

Limitations

The NHIS question asked about individuals' email exchange with clinicians and did not specifically mention secure messaging. Therefore, it is possible that the estimates provided in this study may overestimate secure messaging use among patients. If the NHIS had asked respondents about secure messaging use, it is likely that many would be unfamiliar with the phrase and may not have responded accurately, resulting in underestimation of messaging use. Most email communication is facilitated through patient portals as secure messages, so the estimates presented in this paper around email exchange are likely to be a close approximation of the use of secure messaging.

Not surprisingly, individuals who saw or spoke with a physician in the year preceding the survey had four times greater odds for using email to communicate with their clinicians than individuals who did not have an interaction with a clinician during that time. From these data, it is not clear if the email communication precipitated the clinical encounter or vice versa. There is no way to determine causality with cross-sectional surveys such as NHIS and NAMCS. There was also no information on whether there was bidirectional exchange using email (eg, patient wrote to clinician and received a response). Prior studies on the relationship between secure messaging and reduced health care utilization were mixed, and it is clear from these results that further research is warranted to better understand that relationship [10,11,39,40].

The outcomes of interest and many of the independent variables were based on self-report and therefore subject to recall bias. NCHS, however, uses validated questions that should reduce the effect of that bias.

Missing Data

In 2015, the NAMCS' physician response rate was less than 30%, which could lead to a nonresponse bias. A detailed analysis of potential nonresponse bias conducted by NCHS in 2012 found no evidence of nonresponse bias based on census division, metropolitan practice location, and physician specialty [15]. However, NCHS did find evidence that the sample may overestimate solo practitioners and underestimate in large (11 physicians or greater) and HMO-owned practices. The differences we observed by practice ownership and size (which we proxied as solo vs nonsolo) were large, so even if there was a bias toward the null based on the nonresponse bias identified by NCHS, it is likely that the statistical differences would persist.

We found a larger percentage of missing data for our dependent variable (19%) in the 2014 NAMCS data compared with that in other years, which ranged between 2% and 3%. There is no information in the NAMCS documentation about this gap. It does raise a concern about the viability of the estimates for that survey year, and those data should be viewed with caution. The missing data rates for all other variables across the surveys were less than 5% (see Multimedia Appendix 2).

In contrast, the response rates for the NHIS were acceptable, with a household response rate ranging between 64% and 76%, and a sampled adult response rate of approximately around 80%. The NCHS conducts the survey in-person and uses flash cards to aid in the understanding of the questions. The random selection of the adult for that questionnaire is weighted toward blacks, Asians, and adults older than 65 years to ensure sufficient sample for those populations.

We found a 50% missing rate for the English language variable for 2013. Here again, survey documentation provided no indication about why there was such a difference in missing rates for 2013 (all subsequent years had less than 1% missing), but those data should be viewed with caution. Similarly, 8% of individuals had missing data for the internet use variable in 2015. All other NHIS variables had missing rates below 5%.

Comparison With Prior Work

The overall percentage of almost 17% using email to communicate with their clinicians is lower than what would be expected, given the findings from other studies that indicated that many patients expressed intention to send messages to their clinicians if given the opportunity [5,13,33]. These data use a cross-section of US residents, which includes both healthy and sick individuals. In other studies, patients currently receiving clinical treatment reported few barriers to use, including challenges in accessing the patient portal to send or receive a message, doubts about the reliability of the messaging function or prior bad experiences, concern about imposing on clinicians' time, and perceived resistance to use of messaging among clinical staff [3,49,50]. In addition, studies identified clinician behavior as a significant driver of secure messaging use among patients [14,43]. For example, patients were more likely to initiate their own messages if their clinicians responded quickly, had a higher overall response rate, or initiated their own messages [14]. Our analyses found no difference in access to the capability by physician specialty, even though prior research found most patients sent messages to their primary care clinicians [6]. Therefore, to promote adoption of email communication among patients, it may be important for clinicians to promote and advocate for its use.

Our findings of the characteristics associated with patients' self-reported communication with clinicians were largely consistent with other research conducted at single sites or within large integrated delivery networks. Our study is the first to use a large, nationally representative sample as is available through the NHIS. Consistent with a number of studies [5,13,51,52], we found higher education to be positively associated with secure messaging use. Similar to our findings as well, several studies found privately insured patients were more likely to use secure messaging [13,36,41]. Only one study examined associations with primary language and secure messaging use [33], and their findings were also consistent with our findings on English language familiarity. Similar to our findings, most studies that examined the association by race found that white patients had higher rates of secure messaging use than other races [5,36,41,45,53-55].

After controlling for other characteristics in our multivariate models, we found that women had reduced odds of using email to communicate with clinicians. This is in contrast to a number of studies that found the reverse [5,13,36,41,45,53,55,56]. Our unadjusted numbers showed a higher percentage of female patients reporting email communication with clinicians than males; it was only after adjusting for other sociodemographic factors that we found the reduced odds. The difference between our research and those studies is that our adjusted analyses controlled for education, health insurance, and the English language.

Conclusions

Aligning analyses of the NHIS and NAMCS presents a unique opportunity to understand patients' potential access to and use of secure messaging over time. These analyses are novel because prior studies were based on data from individual health care organizations or integrated delivery networks, or used smaller national survey data that asked about email and internet communication with clinicians. There are no published data that compare secure messaging by practice ownership and size. This is the first study to use a large nationally representative sample to explore the association between selected social determinants of health and email exchange between patients and clinicians.

In 2009, Street et al [31] published a framework that describes how communication functions such as information sharing may directly and indirectly lead to improved patient outcomes. Many secure messaging benefits cited by clinicians—improvements in access, more direct and focused communication between patients and clinic staff, improved efficiency including avoidance of phone tag, improved communication between visits, and improved patient engagement, satisfaction, and trust [3,4]—are constructs in the Street et al [31] pathway between communication functions and improved patient outcomes. Unfortunately, our data demonstrate that there is unequal use of secure messaging among patients based on social and demographic characteristics, and that secure messaging use by patients is not increasing as the functionality is made available through office visits. Developing a better understanding of who is using secure messaging, and whether and how clinicians are encouraging that use, may permit better targeting of secure messaging interventions that encourage secure messaging adoption and use.

Acknowledgments

DG conducted all analysis and wrote and edited all manuscript drafts. HC guided the conceptual development of the research, assisted with analysis, and edited the manuscript. DG's affiliation with The MITRE Corporation is provided for identification purposes only and is not intended to convey or imply MITRE's concurrence with, or support for, the positions, opinions, or viewpoints expressed by the author.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Selected National Health Interview Survey and National Ambulatory Medical Care Survey questions and responses included in analyses: lists the survey questions and response options included in the analyses.
[DOCX File , 25 KB - jmir_v22i5e12611_app1.docx]

Multimedia Appendix 2
Percentage of missing data by survey, variable, and year: provides 2 tables—1 for each survey—with the missing data rates by variable and year for all dependent and independent variables used in the analyses.
[DOCX File , 22 KB - jmir_v22i5e12611_app2.docx]

Multimedia Appendix 3
Prevalence (95% CI) of secure message use by physicians conducting ambulatory care medical visits. Table of the percentages (with 95% CI) of ambulatory medical care visits conducted by a physician who used secure messaging in 2013 and 2016, plus the P value to indicate statistical difference between the years.
[DOCX File , 22 KB - jmir_v22i5e12611_app3.docx]

References

1. Institute of Medicine, Committee on Quality of Health Care in America. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academies Press (US); 2001.
2. Centers for Medicare and Medicaid Services, Department of Health and Human Services. US Government Publishing Office. 2012. Federal Register: Part II - Department of Health and Human Services URL: https://www.gpo.gov/fdsys/pkg/FR-2012-09-04/pdf/2012-21050.pdf [accessed 2018-03-02]

3. Wade-Vuturo AE, Mayberry LS, Osborn CY. Secure messaging and diabetes management: experiences and perspectives of patient portal users. J Am Med Inform Assoc 2013 May 1;20(3):519-525 [FREE Full text] [doi: 10.1136/amiajnl-2012-001253] [Medline: 23242764]
4. Nazi KM. The personal health record paradox: health care professionals' perspectives and the information ecology of personal health record systems in organizational and clinical settings. J Med Internet Res 2013 Apr 4;15(4):e70 [FREE Full text] [doi: 10.2196/jmir.2443] [Medline: 23557596]
5. Haun JN, Patel NR, Lind JD, Antinori N. Large-scale survey findings inform patients' experiences in using secure messaging to engage in patient-provider communication and self-care management: A quantitative assessment. J Med Internet Res 2015 Dec 21;17(12):e282 [FREE Full text] [doi: 10.2196/jmir.5152] [Medline: 26690761]
6. Crotty BH, Tamrat Y, Mostaghimi A, Safran C, Landon BE. Patient-to-physician messaging: volume nearly tripled as more patients joined system, but per capita rate plateaued. Health Aff (Millwood) 2014 Oct;33(10):1817-1822 [FREE Full text] [doi: 10.1377/hlthaff.2013.1145] [Medline: 25288428]
7. Anand SG, Feldman MJ, Geller DS, Bisbee A, Bauchner H. A content analysis of e-mail communication between primary care providers and parents. Pediatrics 2005 May;115(5):1283-1288. [doi: 10.1542/peds.2004-1297] [Medline: 15867036]
8. Heisey-Grove D, Patel V, Searcy T. Health IT. 2015 Oct. Physician Electronic Exchange of Patient Health Information, 2014 URL: https://dashboard.healthit.gov/evaluations/data-briefs/physician-electronic-exchange-patient-health-information.php [accessed 2018-10-23] [WebCite Cache ID 73OWIo390]
9. Henry J, Pylypchuk Y, Patel V. Health IT. 2016 Sep. Electronic Capabilities for Patient Engagement among US Non-Federal Acute Care Hospitals: 2012-2015 URL: https://dashboard.healthit.gov/evaluations/data-briefs/hospitals-patient-engagement-electronic-capabilities-2015.php [accessed 2018-10-23] [WebCite Cache ID 73OWn2jQJ]
10. Liss DT, Reid RJ, Grembowski D, Rutter CM, Ross TR, Fishman PA. Changes in office visit use associated with electronic messaging and telephone encounters among patients with diabetes in the PCMH. Ann Fam Med 2014 Jul;12(4):338-343 [FREE Full text] [doi: 10.1370/afm.1642] [Medline: 25024242]
11. Shimada SL, Hogan TP, Rao SR, Allison JJ, Quill AL, Feng H, et al. Patient-provider secure messaging in VA: variations in adoption and association with urgent care utilization. Med Care 2013 Mar;51(3 Suppl 1):S21-S28. [doi: 10.1097/MLR.0b013e3182780917] [Medline: 23407007]
12. Crotty B, Mostaghimi A, O'Brien J, Bajracharya A, Safran C, Landon B. Prevalence and risk profile of unread messages to patients in a patient web portal. Appl Clin Inform 2015;6(2):375-382 [FREE Full text] [doi: 10.4338/ACI-2015-01-CR-0006] [Medline: 26171082]
13. Tarver WL, Menser T, Hesse BW, Johnson TJ, Beckjord E, Ford EW, et al. Growth dynamics of patient-provider internet communication: Trend analysis using the Health Information National Trends Survey (2003 to 2013). J Med Internet Res 2018 Mar 29;20(3):e109 [FREE Full text] [doi: 10.2196/jmir.7851] [Medline: 29599107]
14. Wolcott V, Agarwal R, Nelson DA. Is provider secure messaging associated with patient messaging behavior? Evidence from the US army. J Med Internet Res 2017 Apr 6;19(4):e103 [FREE Full text] [doi: 10.2196/jmir.6804] [Medline: 28385681]
15. National Center for Health Statistics. Centers for Disease Control and Prevention. 2015. 2015 NAMCS Micro-Data File Documentation URL: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NAMCS/doc2015.pdf [accessed 2018-10-21] [WebCite Cache ID 73Kh6laTL]
16. National Center for Health Statistics. Centers for Disease Control and Prevention. 2014. 2014 NAMCS Micro-Data File Documentation URL: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NAMCS/doc2014.pdf [accessed 2018-10-21] [WebCite Cache ID 73KjebRPS]
17. National Center for Health Statistics. Centers for Disease Control and Prevention. 2011. 2011 NAMCS Micro-Data File Documentation URL: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NAMCS/doc2011.pdf [accessed 2018-10-21] [WebCite Cache ID 73KkRzzPY]
18. National Center for Health Statistics. Centers for Disease Control and Prevention. 2013. 2013 NAMCS Micro-Data File Documentation URL: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NAMCS/doc2013.pdf [accessed 2018-10-21] [WebCite Cache ID 73KjpXDhV]
19. National Center for Health Statistics. Centers for Disease Control and Prevention. 2012. 2012 NAMCS Micro-Data File Documentation URL: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NAMCS/doc2012.pdf [accessed 2018-10-21] [WebCite Cache ID 73KkEWwl6]
20. National Center for Health Statistics. Centers for Disease Control and Prevention. About the Ambulatory Health Care Surveys URL: https://www.cdc.gov/nchs/ahcd/about_ahcd.htm [accessed 2018-10-21] [WebCite Cache ID 73Khg148H]
21. National Center for Health Statistics. Centers for Disease Control and Prevention. 2016. 2016 NAMCS Micro-Data File Documentation URL: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NAMCS/doc2016.pdf [accessed 2019-12-29]
22. National Center for Health Statistics. Centers for Disease Control and Prevention. 2017. 2016 National Health Interview Survey (NHIS) Survey Description URL: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2016/srvydesc.pdf [accessed 2018-10-24] [WebCite Cache ID 73OXTziaL]

23. National Center for Health Statistics. Centers for Disease Control and Prevention. NHIS Data, Questionnaires and Related Documentation URL: https://www.cdc.gov/nchs/nhis/data-questionnaires-documentation.htm [accessed 2018-10-24] [WebCite Cache ID 73OYXDPFB]
24. National Center for Health Statistics. Centers for Disease Control and Prevention. 2016. 2015 National Health Interview Survey (NHIS) Survey Description URL: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2015/srvydesc.pdf [accessed 2018-03-02]
25. National Center for Health Statistics. Centers for Disease Control and Prevention. 2015. 2014 National Health Interview Survey (NHIS) Survey Description URL: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2014/srvydesc.pdf [accessed 2018-03-02]
26. National Center for Health Statistics. Centers for Disease Control and Prevention. 2014. 2013 National Health Interview Survey (NHIS) Survey Description URL: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2013/srvydesc.pdf [accessed 2018-03-02]
27. National Center for Health Statistics. Centers for Disease Control and Prevention. 2013. 2012 National Health Interview Survey (NHIS) Survey Description URL: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2012/srvydesc.pdf [accessed 2018-03-02]
28. National Center for Health Statistics. Centers for Disease Control and Prevention. 2012. 2011 National Health Interview Survey (NHIS) Survey Description URL: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2011/srvydesc.pdf [accessed 2018-03-02]
29. National Center for Health Statistics. Centers for Disease Control and Prevention. 2018. 2017 National Health Interview Survey (NHIS) Survey Description URL: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2017/srvydesc.pdf [accessed 2018-10-24] [WebCite Cache ID 73OXTziaL]
30. National Center for Health Statistics. Centers for Disease Control and Prevention. 2019. 2018 National Health Interview Survey (NHIS) Survey Description URL: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2018/srvydesc.pdf [accessed 2019-12-29]
31. Street RL, Makoul G, Arora NK, Epstein RM. How does communication heal? Pathways linking clinician-patient communication to health outcomes. Patient Educ Couns 2009 Mar;74(3):295-301. [doi: 10.1016/j.pec.2008.11.015] [Medline: 19150199]
32. Mishel MH. Uncertainty in illness. Image J Nurs Sch 1988;20(4):225-232. [doi: 10.1111/j.1547-5069.1988.tb00082.x] [Medline: 3203947]
33. Schickedanz A, Huang D, Lopez A, Cheung E, Lyles CR, Bodenheimer T, et al. Access, interest, and attitudes toward electronic communication for health care among patients in the medical safety net. J Gen Intern Med 2013 Jul;28(7):914-920 [FREE Full text] [doi: 10.1007/s11606-012-2329-5] [Medline: 23423453]
34. Epstein RM, Street RL. Division of Cancer Control and Population Sciences. Bethesda, MD: National Cancer Institute; 2007. Patient-Centered Communication in Cancer Care: Promoting Healing and Reducing Suffering URL: https://cancercontrol.cancer.gov/brp/docs/pcc_monograph.pdf [accessed 2018-11-27]
35. Zhou YY, Kanter MH, Wang JJ, Garrido T. Improved quality at Kaiser Permanente through e-mail between physicians and patients. Health Aff (Millwood) 2010 Jul;29(7):1370-1375. [doi: 10.1377/hlthaff.2010.0048] [Medline: 20606190]
36. Price-Haywood E, Luo Q, Monlezun D. Dose effect of patient-care team communication via secure portal messaging on glucose and blood pressure control. J Am Med Inform Assoc 2018 Jun 1;25(6):702-708. [doi: 10.1093/jamia/ocx161] [Medline: 29444256]
37. Bredfeldt CE, Compton-Phillips AL, Snyder MH. Effects of between visit physician-patient communication on Diabetes Recognition Program scores. Int J Qual Health Care 2011 Dec;23(6):664-673. [doi: 10.1093/intqhc/mzr061] [Medline: 21937586]
38. Schneider M, Howard K. Using technology to enhance discharge teaching and improve coping for patients after stroke. J Neurosci Nurs 2017 Jun;49(3):152-156. [doi: 10.1097/JNN.0000000000000275] [Medline: 28346312]
39. Harris LT, Haneuse SJ, Martin DP, Ralston JD. Diabetes quality of care and outpatient utilization associated with electronic patient-provider messaging: a cross-sectional analysis. Diabetes Care 2009 Jul;32(7):1182-1187 [FREE Full text] [doi: 10.2337/dc08-1771] [Medline: 19366959]
40. Harris LT, Koepsell TD, Haneuse SJ, Martin DP, Ralston JD. Glycemic control associated with secure patient-provider messaging within a shared electronic medical record: a longitudinal analysis. Diabetes Care 2013 Sep;36(9):2726-2733 [FREE Full text] [doi: 10.2337/dc12-2003] [Medline: 23628618]
41. Shimada SL, Allison JJ, Rosen AK, Feng H, Houston TK. Sustained use of patient portal features and improvements in diabetes physiological measures. J Med Internet Res 2016 Jul 1;18(7):e179 [FREE Full text] [doi: 10.2196/jmir.5663] [Medline: 27369696]
42. Chung AE, Sandler RS, Long MD, Ahrens S, Burris JL, Martin CF, et al. Harnessing person-generated health data to accelerate patient-centered outcomes research: the Crohn's and Colitis Foundation of America PCORnet Patient Powered Research Network (CCFA Partners). J Am Med Inform Assoc 2016 May;23(3):485-490 [FREE Full text] [doi: 10.1093/jamia/ocv191] [Medline: 26911821]

43. Alpert JM, Markham MJ, Bjarnadottir RI, Bylund CL. Twenty-first century bedside manner: Exploring patient-centered communication in secure messaging with cancer patients. J Cancer Educ 2019 Jul 24:-. [doi: 10.1007/s13187-019-01592-5] [Medline: 31342283]
44. Goldzweig CL, Orshansky G, Paige NM, Towfigh AA, Haggstrom DA, Miake-Lye I, et al. Systematic Review: Secure Messaging Between Providers and Patients, and Patients' Access to Their Own Medical Record: Evidence on Health Outcomes, Satisfaction, Efficiency and Attitudes. Washington (DC): Department of Veterans Affairs (US); 2012.
45. North F, Crane SJ, Stroebel RJ, Cha SS, Edell ES, Tulledge-Scheitel SM. Patient-generated secure messages and eVisits on a patient portal: are patients at risk? J Am Med Inform Assoc 2013;20(6):1143-1149 [FREE Full text] [doi: 10.1136/amiajnl-2012-001208] [Medline: 23703826]
46. Heyworth L, Clark J, Marcello TB, Paquin AM, Stewart M, Archambeault C, et al. Aligning medication reconciliation and secure messaging: qualitative study of primary care providers' perspectives. J Med Internet Res 2013 Dec 2;15(12):e264 [FREE Full text] [doi: 10.2196/jmir.2793] [Medline: 24297865]
47. Ozkaynak M, Johnson S, Shimada S, Petrakis BA, Tulu B, Archambeault C, et al. Examining the multi-level fit between work and technology in a secure messaging implementation. AMIA Annu Symp Proc 2014;2014:954-962 [FREE Full text] [Medline: 25954403]
48. Wooldridge AR, Carayon P, Hoonakker P, Musa A, Bain P. Technology-mediated communication between patients and primary care clinicians and staff. Proc Hum Factors Ergon Soc Annu Meet 2016;60(1):556-560. [doi: 10.1177/1541931213601128]
49. Haun JN, Lind JD, Shimada SL, Martin TL, Gosline RM, Antinori N, et al. Evaluating user experiences of the secure messaging tool on the Veterans Affairs' patient portal system. J Med Internet Res 2014 Mar 6;16(3):e75 [FREE Full text] [doi: 10.2196/jmir.2976] [Medline: 24610454]
50. Lam R, Lin VS, Senelick WS, Tran HP, Moore AA, Koretz B. Older adult consumers' attitudes and preferences on electronic patient-physician messaging. Am J Manag Care 2013 Nov;19(10 Spec No):eSP7-eS11 [FREE Full text] [Medline: 24511886]
51. White CB, Moyer CA, Stern DT, Katz SJ. A content analysis of e-mail communication between patients and their providers: patients get the message. J Am Med Inform Assoc 2004;11(4):260-267 [FREE Full text] [doi: 10.1197/jamia.M1445] [Medline: 15064295]
52. Graetz I, Gordon N, Fung V, Hamity C, Reed ME. The digital divide and patient portals: internet access explained differences in patient portal use for secure messaging by age, race, and income. Med Care 2016 Aug;54(8):772-779. [doi: 10.1097/MLR.0000000000000560] [Medline: 27314262]
53. Cronin RM, Davis SE, Shenson JA, Chen Q, Rosenbloom ST, Jackson GP. Growth of secure messaging through a patient portal as a form of outpatient interaction across clinical specialties. Appl Clin Inform 2015;6(2):288-304 [FREE Full text] [doi: 10.4338/ACI-2014-12-RA-0117] [Medline: 26171076]
54. Masterman M, Cronin RM, Davis SE, Shenson JA, Jackson GP. Adoption of secure messaging in a patient portal across pediatric specialties. AMIA Annu Symp Proc 2016;2016:1930-1939 [FREE Full text] [Medline: 28269952]
55. Chung S, Panattoni L, Chi J, Palaniappan L. Can secure patient-provider messaging improve diabetes care? Diabetes Care 2017 Oct;40(10):1342-1348. [doi: 10.2337/dc17-0140] [Medline: 28807977]
56. Reed M, Graetz I, Gordon N, Fung V. Patient-initiated e-mails to providers: associations with out-of-pocket visit costs, and impact on care-seeking and health. Am J Manag Care 2015 Dec 1;21(12):e632-e639 [FREE Full text] [Medline: 26760425]

Abbreviations

EHR: electronic health record
HMO: Health Maintenance Organization
IOM: Institute of Medicine
IT: information technology
MACRA: Medicare Access and CHIP Reauthorization Act of 2015
NAMCS: National Ambulatory Medical Care Survey
NCHS: National Center for Health Statistics
NHIS: National Health Interview Survey

Edited by G Eysenbach; submitted 25.10.18; peer-reviewed by D Liss, L Mayberry; comments to author 31.03.19; revised version received 31.12.19; accepted 27.01.20; published 01.05.20.

Please cite as:
Heisey-Grove DM, Carretta HJ
Disparities in Secure Messaging Uptake Between Patients and Physicians: Longitudinal Analysis of Two National Cross-Sectional Surveys
J Med Internet Res 2020;22(5):e12611
URL: https://www.jmir.org/2020/5/e12611
doi:10.2196/12611
PMID:32356775

Original Paper

Secure Asynchronous Communication Between Smokers and Tobacco Treatment Specialists: Secondary Analysis of a Web-Assisted Tobacco Intervention in the QUIT-PRIMO and National Dental PBRN Networks

Rajani Shankar Sadasivam[1*], PhD; Ariana Kamberi[1*], MBA; Kathryn DeLaughter[2*], MA; Barrett Phillips[3*], MPH; Jessica H Williams[4*], MPH, PhD; Sarah L Cutrona[1*], MD; Midge N Ray[4*], PhD; Gregg H Gilbert[4*], DDS, MBA; Thomas K Houston[1*], MD, MPH; QUITPRIMO[4*]; National Dental PBRN Collaborative Group[4*]

[1]University of Massachusetts Medical School, Worcester, MA, United States

[2]Edith Nourse Rogers Memorial Veterans Hospital, Bedford, MA, United States

[3]Veterans Affairs Central Western Massachusetts Healthcare System, Leeds, MA, United States

[4]University of Alabama at Birmingham, Birmingham, AL, United States

[*]all authors contributed equally

Corresponding Author:
Rajani Shankar Sadasivam, PhD
University of Massachusetts Medical School
365 Plantation Street
Worcester, MA, 01605
United States
Phone: 1 508 856 8923
Email: rajani.sadasivam@umassmed.edu

Abstract

Background: Within a web-assisted tobacco intervention, we provided a function for smokers to asynchronously communicate with a trained tobacco treatment specialist (TTS). Previous studies have not attempted to isolate the effect of asynchronous counseling on smoking cessation.

Objective: This study aimed to conduct a semiquantitative analysis of TTS-smoker communication and evaluate its association with smoking cessation.

Methods: We conducted a secondary analysis of data on secure asynchronous communication between trained TTSs and a cohort of smokers during a 6-month period. Smokers were able to select their preferred TTS and message them using a secure web-based form. To evaluate whether the TTS used evidence-based practices, we coded messages using the Motivational Interviewing Self-Evaluation Checklist and Smoking Cessation Counseling (SCC) Scale. We assessed the content of messages initiated by the smokers by creating topical content codes. At 6 months, we assessed the association between smoking cessation and the amount of TTS use and created a multivariable model adjusting for demographic characteristics and smoking characteristics at baseline.

Results: Of the 725 smokers offered asynchronous counseling support, 33.8% (245/725) messaged the TTS at least once. A total of 1082 messages (TTSs: 565; smokers 517) were exchanged between the smokers and TTSs. The majority of motivational interviewing codes were those that supported client strengths (280/517, 54.1%) and promoted engagement (280/517, 54.1%). SCC code analysis showed that the TTS provided assistance to smokers if they were willing to quit (247/517, 47.8%) and helped smokers prepare to quit (206/517, 39.8%) and anticipate barriers (197/517, 38.1%). The majority of smokers' messages discussed motivations to quit (234/565, 41.4%) and current and past treatments (talking about their previous use of nicotine replacement therapy and medications; 201/565, 35.6%). The majority of TTS messages used behavioral strategies (233/517, 45.1%), offered advice on treatments (189/517, 36.5%), and highlighted motivations to quit (171/517, 33.1%). There was no association between the amount of TTS use and cessation. In the multivariable model, after adjusting for gender, age, race, education, readiness at baseline, number of cigarettes smoked per day at baseline, and the selected TTS, smokers messaging the TTS one or two times had a smoking cessation odds ratio (OR) of 0.8 (95% CI 0.4-1.4), and those that messaged the TTS more than two times had a smoking cessation OR of 1.0 (95% CI 0.4-2.3).

Conclusions: Our study demonstrated the feasibility of using asynchronous counseling to deliver evidence-based counseling. Low participant engagement or a lack of power could be potential explanations for the nonassociation with smoking cessation. Future trials should explore approaches to increase participant engagement and test asynchronous counseling in combination with other approaches for improving the rates of smoking cessation.

(J Med Internet Res 2020;22(5):e13289) doi:10.2196/13289

KEYWORDS

distance counseling; tobacco cessation; internet-based intervention; smoking cessation

Introduction

Background

Smoking is the number one preventable cause of death [1]. Nearly 1 in 5 adults in the United States, or 42.1 million people, currently smoke cigarettes [2]. Quitting reduces the risk of smoking-related diseases, including cardiovascular diseases [3]. Current guidelines recommend that smokers use behavioral counseling to aid their quitting efforts [4]. However, most smokers do not use counseling in their quit attempts. For instance, despite the wide availability of Quitlines, a telephone-based tobacco counseling that has been established in all 50 states, it has been reported that only 1% to 2% of smokers use Quitline services [5,6]. More innovations are needed to increase access to counseling resources [1].

Providing behavioral counseling by trained tobacco treatment specialists (TTSs) over the internet may help overcome barriers to access. Internet access is almost ubiquitous in the United States, especially with the increasing use of mobile phones [7-9]. Users have also become comfortable communicating over the web (eg, emails, text messages, and on social networks). Web-based counseling is usually delivered synchronously through live chat, where the counselor and smoker are both online and engage via video, audio, or messaging. Synchronous counseling has the benefit of the smokers being able to directly interact with the counselors, and for the counselors to be able to tailor the counseling to the smoker. Previous papers have evaluated synchronous counseling [10-12]. However, synchronous counseling may not be convenient for all smokers as they need to coordinate their schedules with the counselors. Asynchronous counseling that allows the smokers the convenience of communicating with the counselor at any time can be used to augment synchronous counseling. Asynchronous counseling can be achieved using a secure Health Insurance Portability and Accountability Act–compliant form of email or secure messaging. Previous studies have provided functions to communicate asynchronously with experts such as a question and answer email service or a text messaging interface for communicating with a human coach [13-15]. These functions have mostly been provided within the context of a larger intervention. To our knowledge, none of these studies conducted secondary analysis to understand and explore the effect of these counseling efforts.

Objective

Our objective is to conduct a secondary evaluation of asynchronous web-based counseling provided by trained TTSs to smokers participating in the QUIT-PRIMO randomized trial testing a web-assisted tobacco intervention (WATI) [16,17]. As published before [16-18] and further described below, in the QUIT-PRIMO randomized trial, the recruited smokers were randomized to the following three groups: a WATI control (control), the WATI enhanced with proactive, pushed tailored email motivational messaging (messaging), and the WATI with messaging further enhanced with personal secure messaging with a TTS (personalized). Although we found significant differences between those who received the motivational messages (personalized and messaging groups) versus the control (17%; odds ratio [OR] 1.69; 95% CI 1.03-2.8; *P*=.04), we found no differences between the personalized and the messaging group of the study.

Understanding TTS use by the smokers might provide us more insight into why the access to TTS did not result in an increase in smoking cessation in our study and may help in designing more patient-centered asynchronous counseling support in future web-based interventions. In this paper, we add to this literature by conducting a semiquantitative analysis of the TTS-smoker communication and evaluating its association with smoking cessation. We report the frequency of TTS-smoker communication, whether the TTSs were able to incorporate evidence-based practices, such as motivational interviewing, into their counseling. Our analysis also included the communication initiated by the smokers.

Methods

Study Design, Setting, and Sample

We conducted a longitudinal, observational cohort study of secure asynchronous communication between 3 trained TTSs and 245 smokers. The communication occurred during a 6-month period [16]. For the purpose of this study, we only included smokers from the personalized group. These smokers were able to select from among 3 TTSs and initiate the communication by messaging their TTS using a secure web-based form. The TTSs were, then, able to message back using an administrative portal. We implemented the *secure* communication similar to how it is implemented on web-based patient portals. When someone initiated a contact through the website, an email alert was sent to the smoker and the TTS. This email alert did not contain any personal identifiers and the smoker and the TTS had to log in to the website to view their messages. Our website was available only through an encrypted channel using the hypertext transfer protocol secure.

The approach for selecting and messaging the TTS and coding messages is described in detail below (Choosing a Tobacco Treatment Specialist in the Personalized Group). Our study was

approved by the University of Alabama at Birmingham and the University of Massachusetts Medical School Institutional Review Boards.

The QUIT-PRIMO Randomized Trial

Recruitment

Smokers were recruited during June 2010-March 2012. Smoker recruitment was conducted through medical and dental practice patient referrals as well as through Google advertisements [16-19]. Our recruitment process has been described in detail in our previous publications. Briefly, primary care medical practices were recruited from a registered database of internal medicine and family or general practitioners. Dental practices were recruited from state lists of registered dentists and through the Dental Practice–Based Research Network. At these practices, we implemented an electronic referral program that allowed providers (doctors, nurses, or other clinical staff) to recruit smokers to the Decide2Quit.org WATI at the point of care by entering their email addresses into a web-based form. When smokers were electronically referred, they were sent up to 10 email messages encouraging registration over an 8-week period or until the patient registered. To participate on Decide2Quit.org, all smokers completed a web-based consent and registered on the website. To recruit smokers through Google advertisements, 3 advertisements were posted on Google AdWords. Advertisements were linked to searches for keywords related to smoking (eg, smoking, quit smoking, stop smoking, quit, quit smoking tips, and quit smoking programs) and included a link that took participants directly to the Decide2Quit.org home page where they could choose to register as new participants. Each smoker had access to Decide2quit.org for 6 months.

Design of the Randomized Trial

The Decide2Quit.org WATI was designed for all smokers, supporting cessation induction for those not ready to quit and acting as an aid to cessation for those preparing to quit [16-19]. Decide2Quit.org was implemented as an adaptable service with modules that could be engaged based on assigned group. Thus, smokers were allocated to one of three increasingly enhanced versions of Decide2Quit.org. The interventions received by the three randomized groups of smokers are described below.

The Control Group

Smokers randomized to the active control received *an interactive, tailored quit smoking website*. This module included motivational information tailored to readiness to quit (not thinking of quitting, thinking of quitting, and preparing to quit), and interactive risk, decisional balance, and cessation barrier calculators, and games linking the chemicals in smoking with their other uses (eg, formaldehyde is used in both cigarettes and in embalming). The control also included a library of informational resources about smoking, and sections on seeking social support and talking to your doctor about quitting. These tools are variations of those in many previously studied evidence-based WATI.

The Messaging Group

Pushed Motivational Email Messages Module Plus Control Module

For this group, we enhanced the control with a pushed motivational email messaging system. Brief motivational email messages were further tailored to an individual smoker's readiness to quit (not ready to quit, thinking about quitting, preparing to quit, and actively quitting). In the first week of registration, 4 email messages were sent to the smoker, followed thereafter by 2 email messages per week. To enhance the personal relevance of messages, our motivational email messaging system included messages written by smokers for other smokers [20].

Personalized Group: Personal Support Module, Messages Module, and Control Module

In addition to the above functions, the personal support module further included an innovative secure messaging portal allowing asynchronous electronic communication between smokers and trained TTSs hired to participate in the intervention team. Further details on how the patient chose and communicated with the TTS are provided below.

Choosing a Tobacco Treatment Specialist in the Personalized Group

During registration over the web, smokers were asked to select between 3 trained TTSs. Smokers were presented the counseling philosophies of the 3 trained TTSs to facilitate this selection. The 3 TTSs created their philosophies from personal experiences and accredited TTS training [21]. Of the 3 TTSs, 2 were trained at the University of Mississippi, Medical Center ACT, while 1 was trained at the University of Massachusetts, Center for Tobacco Treatment Research and Training. Both training programs were accredited by the Council for Tobacco Treatment Training Programs **[22,23].** Content of this training consisted of bio, psycho, and social determinants of nicotine dependence, pharmacotherapy, and counseling theory and practice, which included motivational interviewing techniques, treatment strategies, and program and system issues.

Counseling philosophies of the 3 TTSs were as follows:

> *It is not about the quantity of your life, it is about the quality. Smoking harms nearly every organ in the body and one out of every two smokers will die prematurely from a smoking-related disease. When you think about it, those are not very good odds. Many of my loved ones struggle with tobacco use so I know first-hand how difficult quitting can be. This is why I take a holistic approach to helping you quit tobacco. It is important for you to understand ALL the reasons why you smoke and for you to think about strategies that will work best for you—I want to help you do that. Quitting tobacco may be one of the most difficult things you will do for yourself but is also one of the most important. Time is ticking...what are you waiting for?* [TTS 1]
>
> *Being healthy and happy is about learning what's important to you and what you are willing to do to*

> *make that happen. Quitting smoking is no different. For some it's as simple as wanting more time with their children or grandchildren, for others perhaps it's about saving money to go on a trip. Either way, small steps now can make a big difference in years to come. For me, I like to help people look at why they want to quit and the reasons they smoke to figure out the best plan for them. Each person is unique and will have their own reasons for quitting smoking. By learning what motivates you and why you do certain things, you are better prepared to make healthy lifestyle changes. This way the skills learned from quitting smoking can be used in other areas of life.* [TTS 2]
>
> *Quitting smoking can truly be one of the biggest challenges you may face...As a prior smoker, I know first-hand the hold nicotine, a highly addictive drug, can have on one's life. Quitting enabled me to see how much life I was missing. I would like to help you quit. Whether you are ready to quit today or just barely considering quitting, there are many things we can talk about! It is important for each smoker to know their own motivations. One very important motivation is to quit for those around you. Lead by example; kids of parents who quit while they are young are 5 times less likely to smoke themselves, and kids with parents who successfully quit are three times as likely to successfully quit themselves. I can help you to stop being "the smoker"; get rid of the shadow and join the ranks of those proud to say they are "smoke-free." Choosing to quit today can mean a new life tomorrow.* [TTS 3]

Smokers were given the option to send a message at the time of registration or proceed to the website and send a message later. After login, smokers were presented with a web form (similar to an email form on a website) that included a textbox for writing the subject and another for writing the message to TTS (Multimedia Appendix 1). Please note that we did not require smokers to send messages. Once they submitted the message, an email was sent to the participant confirming their submission and indicating that the TTS will respond in 2 business days. The TTSs logged into the administrative portal daily to view any messages they had received. Each message on the administrative portal listed the assigned TTS. Messages were typically replied to by the TTSs within 2 business days. Once the TTS responded, the smokers were sent an email notifying that the TTS had responded to their message. Smokers had to login to the system to read the message. We designed the display of the message to resemble an email inbox. New messages were indicated by the subject in bold font and the smoker had to click on the message subject to read the response.

Coding Process

To code messages and collect codes by categories, we used MAXQDA 10 (VERBI Software) [24-26], a qualitative data analysis software. This software facilitates coding by creating a hierarchical coding dictionary that can be directly mapped to the messages within the software. Authors AK, KL, BP, JW, HC, and SC were all involved in coding the messages. The principal investigator TH guided the process and trained the initial set of coders (BP and AK). BP participated in a training offered by the MAXQDA company on the tool and trained the rest of the coders on the tool. They also created a documentation of codes and coded an initial set of 10 messages under the guidance of TH. Each additional coder was integrated using the following process: They were provided the documentation of codes and asked to code the same 10 messages that had originally been coded by the first two coders (BP and AK). We, then, compared the new and original codes for interrater reliability (IRR). We reviewed the IRR results and compared them with the coding completed by the original coders. On the basis of this comparison and discussion, changes were made to code definitions. Each message was coded by at least 2 coders, and we calculated IRR for each set of 100 messages. After discussion, coder pairs reached agreement on codes used and resolved any discrepancies during the coding process. In the end, 100% agreement was achieved by consensus.

Development of Coding Schema for Analyzing the Tobacco Treatment Specialist Messages

Our coding schema was developed using the Motivational Interviewing Self-Evaluation Checklist and the Smoking Cessation Counseling (SCC) Scale. Motivational interviewing is a directive**,** patient-centered counseling style for eliciting behavior change by helping people to explore, clarify, and resolve ambivalence [27]. The Motivational Interviewing Self-Evaluation Checklist is a 7-item checklist, used as a self-evaluation tool to improve motivational interviewing skills [28-32]. This checklist is focused on engagement, assessing motivation, addressing ambivalence, promoting internal motivation, eliciting change talk, rolling with resistance, and supporting client strength. The SCC Scale is a 24-item questionnaire based on a 4-level response format. It was originally developed for nurses by the US Department of Health and Human Services to help in assessing, improving, or testing evidence-based methodologies for SCC [33]. The scale has both clinical relevance and research applications and can be used clinically to assess the quality of smoking cessation services [33,34]. We revised the SCC by converting its 4-level response to a 2-level response format, and by using the 17 items that focused on advanced and basic counseling after revising the scale.

We also assessed the content of messages initiated by the smokers. For this purpose, we iteratively developed a topical content coding schema (Multimedia Appendix 2). To develop this coding system, we first reviewed 20 messages for both TTSs and smokers and developed a preliminary coding schema based on these. We used this schema to code 20 additional messages, and then refined and added new codes if needed. We proceeded in blocks of 20 messages, and code definitions were revised several times before we finalized our content code category. The content themes included website content; treatment questions: general advice; treatment questions: advice behavioral; treatment questions: advice on treatments and over the counter medications (Rx and OTC); motivations; sociocultural; health; and current or past treatment and feedback from patients.

Data Collection

Data were collected online on the WATI. During registration, we collected demographic and smoking characteristics, and readiness to quit. The TTS and smoker messages were also recorded in the database. At follow-up, the 30-day point prevalence smoking cessation rate was assessed as self-report using the question *Did you smoke any cigarettes during the past 30 days?* [35]. We also assessed on a Likert scale whether communicating with the TTS was helpful to the smoker.

Data Analysis

Our analysis was conducted using STATA version 12 (StataCorp). All analyses were two-sided and alpha error was set at .05. We compared demographic characteristics and smoking behavior between those participants who had sent at least one message to the TTS and those who did not using the chi-square test or the Mantel-Haenszel method (MH odds) for trend. We used descriptive statistics (means or medians) to assess the codes in the TTS messages (Motivational Interviewing Self-Evaluation Checklist and SCC Scale codes). We compared the topical content codes in the TTS messages to the codes in the smoker messages. At the smoker level, we assessed the association between 6-month cessation and the frequency of each code in the TTS messages received by the smoker using separate logistic regression models for each code. For the independent variable for this analysis, we used the number of times the code appeared in messages to best represent the variations in TTS-smoker communication between the smokers. We assessed the association between smoking cessation and the selected TTS by adjusting for demographic characteristics and smoking characteristics at baseline. Furthermore, we assessed the association between 6-month cessation and the amount of TTS use and created a multivariable logistic model; adjusting for demographic characteristics, smoking characteristics at baseline, and the selected TTS. For this analysis, our dependent variable was the 30-day point prevalence cessation rate as described above. All those with missing follow-up data were coded as continued smoking. The amount of TTS use was the total number of messages sent to the TTS by each smoker. For our analysis, we categorized the amount of TTS use as follows: did not message the TTS, messaged the TTS 1 or 2 times, and messaged the TTS more than 2 times. We implemented a selection model using inverse probability weighting to determine the potential effect of the missing data [36,37]. First, based on covariates available within the dataset, we developed a logistic regression model to predict the amount of missing data. Then, we calculated the inverse probability of not being missing and weighted the main analysis by this probability.

Results

Comparing Characteristics of Smokers Using a Tobacco Treatment Specialist and Those Who Did Not Use a Tobacco Treatment Specialist

Approximately, a third of smokers (245/725, 33.8%) who were offered asynchronous counseling support sent messages at least once to the TTS (see the example of communication between smokers and TTSs in Multimedia Appendix 3). Those smokers who allowed smoking at home and who had visited a WATI before this study were more likely to use the TTS (P=.04 for both; Table 1). Among those who used the TTS and completed follow-up, 71% (32/45) reported that the communication was very helpful or somewhat helpful. In all, 29% (13/45) indicated that the communication was not very helpful or not at all helpful.

Table 1. Demographics and smoking behavior of participants, comparing those who messaged the tobacco treatment specialist (TTS) at least once and those who did not.

Patient characteristics	Total smokers included in the study (N=725), n	Messaged the TTS at least once during the 6-month period (n=245), n (%)	Did not message the TTS even once during the 6-month period, (n=480), n (%)	*P* value
Overall	725	245 (33.8)	480 (66.2)	N/A[a]
Patient sex				**.28**
Female	469	165 (35.2)	304 (64.8)	N/A
Male	256	80 (31.2)	176 (68.8)	N/A
Patient age (years)				**.05[b]**
19-34	151	37 (24.5)	114 (75.5)	N/A
35-55	361	127 (35.2)	234 (64.8)	N/A
55-64	167	63 (37.7)	104 (62.3)	N/A
≥65	46	18 (39.1)	28 (60.9)	N/A
Patient race				**.06**
White	605	195 (32.2)	410 (67.8)	N/A
Black or African American	60	28 (46.7)	32 (53.3)	N/A
Other	22	9 (40.9)	13 (59.1)	N/A
Patient education				**.09[b]**
Some high school and high school graduate	245	71 (29.0)	174 (71.0)	N/A
Some college	306	109 (35.6)	197 (64.4)	N/A
College graduate or more	166	64 (38.5)	102 (61.5)	N/A
Readiness to quit				**.96[b]**
Not thinking of quitting and thinking of quitting	518	176 (34.0)	342 (66.0)	N/A
Set a quit date and already quit	202	69 (34.2)	133 (65.8)	N/A
Allow smoking at home				**.04**
No	408	125 (30.6)	283 (69.4)	N/A
Yes	317	120 (37.8)	197 (62.2)	N/A
Number of cigarettes smoked per day				**.67[b]**
0-10	218	72 (33.0)	146 (67.0)	N/A
11-20	355	117 (32.9)	238 (67.1)	N/A
>20	152	56 (36.8)	96 (63.2)	N/A
Visited other smoking cessation websites before the study				**.04**
No	565	180 (31.8)	385 (68.2)	N/A
Yes	160	65 (40.6)	95 (59.4)	N/A
Quit attempt (1 day or more) in the past 12 months				**.70**
No	321	106 (33.0)	215 (67.0)	N/A
Yes	404	139 (34.4)	265 (65.6)	N/A

[a]N/A: not applicable.

[b]Test for trend analysis.

Message Use Frequency and Volume

A total of 1082 messages (TTSs: 565/1082, 52.2%; smokers: 517/1082, 47.8%) were exchanged between the smokers and TTSs. Among those smokers who messaged at least once, the mean number of messages sent was 2.6 (SD 4.5, median 1, IQR 1). The frequency of messages sent by TTS 1 was 341 (mean 6.1, SD 6.4; median 4, IQR 6), by TTS 2 was 370 (mean 9.7, SD 12.2; median 3, IQR 11), and by TTS 3 was 371 (mean 6.7, SD 8.9; median 2, IQR 7).

Comparison of Tobacco Treatment Specialist Motivational Messaging Codes

Motivational Interviewing Self-Evaluation Checklist

The majority of motivational interviewing codes were those that supported client strengths (280/517, 54.1%) and promoted engagement (280/5170, 54.1%; Table 2). For example, a TTS message highlighting the client strength was **“**well it sounds like you certainly have good experience with nicotine replacement, which is good because you have an idea of what works and what doesn’t.”

Table 2. Frequency of codes in messages (N=517).

Scale, category of codes, and codes	Messages, n (%)
Motivational Interviewing Scale	
Supporting client strengths	**280 (54.1)**
Explored previous successes	N/A[a]
Explored positive qualities	N/A
Accentuated *any* motivation for change	N/A
Highlighted any efforts towards change	N/A
Used affirmations to highlight strengths, motivation	N/A
Engagement	**280 (54.1)**
Worked to fully understand the problem and the client's perspective before moving toward change	N/A
Focused on engagement before change	N/A
Used reflective listening to convey empathy and understanding	N/A
Used affirmations to build a positive relationship	N/A
Eliciting change talk	**110 (21.3)**
Asked about concerns using open-ended questions or reflective listening	N/A
Asked for elaboration about concerns	N/A
Explored client values as they relate to change	N/A
Selectively responded to change talk with curiosity, interest	N/A
Explored pros and cons or decisional balance	N/A
Used *low-threshold* questions	N/A
Promoting internal motivation	**106 (20.5)**
Elicited and respected the client's goals for treatment	N/A
Explored values underlying the motivation for change	N/A
Supported autonomy in decision making	N/A
Assessing motivation	**78 (15.1)**
Identified a target behavior	N/A
Identified stage of change	N/A
Used importance, confidence, and readiness ruler	N/A
Differentiated between different areas of motivation (eg, substance use vs mental health; treatment vs change)	N/A
Addressing ambivalence	**72 (13.9)**
Normalized ambivalence	N/A
Explored ambivalence	N/A
Reframed ambivalence	N/A
Used a decisional balance	N/A
Avoided direct persuasion	N/A
Explored pros and cons of change	N/A
Rolling with resistance	**48 (9.3)**
Affirmation (external reframe)	N/A
Empathic response	N/A
Reflective listening	N/A
Providing choice	N/A
Nondefensive response	N/A
Smoking Cessation Counseling Scale	

Scale, category of codes, and codes	Messages, n (%)
Basic counseling	
If willing to quit, provide assistance	247 (47.8)
Ask if willing to quit	26 (5.0)
If not quitting, help identify barriers	23 (4.4)
Advanced counseling	
Helping smokers prepare to quit[b]	206 (39.8)
Helping smokers anticipate barriers[c]	197 (38.1)
Managing relapses[d]	73 (14.1)
Advising smokers to set a quit date	47 (9.2)
Getting social support[e]	46 (8.9)

[a]N/A: not applicable.

[b]Helping smokers prepare to quit codes include identify reasons and benefits for quitting; provide information for follow-up visits with doctor; and recommend over-the-counter nicotine patch, other medications.

[c]Help smokers anticipate barriers codes include anticipate challenges in the beginning; anticipate nicotine withdrawal; alcohol is associated with relapse; and review past quit attempts, what helped, and what led to relapse.

[d]Manage relapses codes include if relapse occurs, review and learn from experience; if relapse occurs, repeat quit attempts; if relapse occurs, reassess problems; and total abstinence is essential.

[e]Getting social support codes include get support from family, friends, and coworkers; and other smokers in house are not helpful.

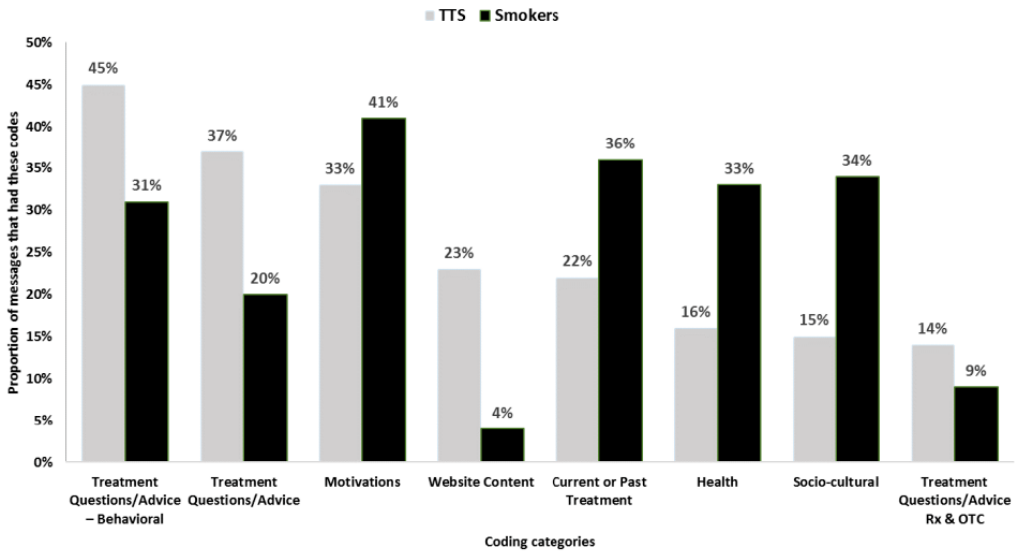
Smoking Cessation Counseling Scale

SCC Scale code analysis showed that the TTS messages provided assistance to smokers if they were willing to quit (247/517, 47.8%; Table 2). For example, a TTS message that included this code was *When you feel you are ready, I am here to help you.* The TTS messages also included the use of advanced counseling strategies to help the smokers prepare to quit (206/517, 39.8%) and anticipate barriers (197/517, 38.1%), such as *Have you thought about ways other than smoking that can help you deal with stress after you quit? Perhaps this is an important first step. It helps to plan as much as possible before you quit to avoid slips*.

Topical Content Codes: Comparison of Tobacco Treatment Specialist and Smoker Messages

TTS messages and smoker messages differed in their content (Figure 1). The majority of smoker messages discussed motivations to quit (234/565, 41.4%) and current and past treatments, including their previous use of nicotine replacement therapy and medications (201/565, 35.6%). An example smoker message is as follows:

Figure 1. Proportion of topical content in messages: Smoker and tobacco treatment specialist messages. TTS: tobacco treatment specialist.



I just lost my mom and step father, both from smoking related disease's. I don't want to die like that ,so i will use all help, and support i can get.

The majority of TTS messages used behavioral strategies (233/517, 45.1%), offered advice on treatments (189/517, 36.5%), and highlighted motivations to quit (171/517, 33.1%).

In the following example, a TTS response established the importance of general behavioral strategy:

> *Most importantly, you should ask your husband to not smoke around you. Your friends and family can be a big help during your quit attempt if you let them know what you need and they support you in quitting. Talk to them about what you will be experiencing during your quit attempt. For example, you may find you have a shorter temper during your quit process. Let your friends or family know this and not to take it personal if you are grumpy.*

Association Between the Point Prevalence Cessation Rate Assessed at 6 Months Using a 30-Day Window and the Selected Tobacco Treatment Specialist, the Content of Tobacco Treatment Specialist Motivational Messages, and Amount of Tobacco Treatment Specialist Use

After adjusting for gender, age, race, education, readiness at baseline, and number of cigarettes smoked per day at baseline, compared with TTS 1, the odds of quitting were lower for TTS 2 (OR 0.4; 95% CI 0.1-1.3) and TTS 3 (OR 0.9 95% CI 0.4-2.4), but not significant. There was no association between the codes in the TTS messages and smoking cessation (Table 3). There was also no association between the total SCC score and smoking cessation (OR 1.02; 95% CI 0.96-1.08).

The amount of TTS use (ie, the total number of messages sent to the TTS by each smoker) ranged from 0 to 42. In all, 66.2% (480/725) of smokers had never used the TTS, 26.1% (189/725) of smokers had used the TTS one or two times, and 7.7% (56/725) of smokers had used it more than two times. There was no association between the amount of TTS use and cessation. In the multivariable model after adjusting for gender, age, race, education, readiness at baseline, number of cigarettes smoked per day at baseline, and the selected TTS, compared with smokers who did not message the TTS, smokers messaging the TTS one or two times had a smoking cessation OR of 0.8 (95% CI 0.4-1.4), and those that messaged the TTS more than two times had a smoking cessation OR of 1.0 (95% CI 0.4-2.3; P=.50 for trend across categories). In the inverse probability weighted model, the smokers messaging the TTS one or two times had a smoking cessation OR of 0.7 (95% CI 0.4-1.3), and those that messaged the TTS more than two times had a smoking cessation OR of 0.7 (95% CI 0.3-1.7; P=.20 for trend across categories).

Table 3. Association between smoking cessation and the frequency of codes.

Scale, category of codes, and codes	Odds ratio (95% CI) using separate logistic regression models for each code
Motivational Interviewing Scale	
Supporting client strengths	**1.12 (0.98-1.32)**
Explored previous successes	N/A[a]
Explored positive qualities	N/A
Accentuated *any* motivation for change	N/A
Highlighted any efforts towards change	N/A
Used affirmations to highlight strengths, motivation	N/A
Engagement	**1.06 (0.88-1.27)**
Worked to fully understand the problem and the client's perspective before moving toward change	N/A
Focused on engagement before change	N/A
Used reflective listening to convey empathy and understanding	N/A
Used affirmations to build a positive relationship	N/A
Eliciting change talk	1.10 (0.74-1.62)
Asked about concerns using open-ended questions or reflective listening	N/A
Asked for elaboration about concerns	N/A
Explored client values as they relate to change	N/A
Selectively responded to change talk with curiosity, interest	N/A
Explored pros and cons or decisional balance	N/A
Used *low-threshold* questions	N/A
Promoting internal motivation	**1.20 (0.74-1.82)**
Elicited and respected the client's goals for treatment	N/A
Explored values underlying the motivation for change	N/A
Supported autonomy in decision making	N/A
Assessing motivation	**1.05 (0.52-2.09)**
Identified a target behavior	N/A
Identified stage of change	N/A
Used importance, confidence, and readiness ruler	N/A
Differentiated between different areas of motivation (eg, substance use vs mental health; treatment vs change)	N/A
Addressing ambivalence	**0.94 (0.55-1.63)**
Normalized ambivalence	N/A
Explored ambivalence	N/A
Reframed ambivalence	N/A
Used a decisional balance	N/A
Avoided direct persuasion	N/A
Explored pros and cons of change	N/A
Rolling with resistance	**0.82 (0.35-1.92)**
Affirmation (external reframe)	N/A
Empathic response	N/A
Reflective listening	N/A
Providing choice	N/A

Scale, category of codes, and codes	Odds ratio (95% CI) using separate logistic regression models for each code
Nondefensive response	N/A
Smoking Cessation Counseling Scale	
Basic counseling	
If willing to quit, provide assistance	1.0 (0.69-1.42)
Ask if willing to quit	0.72 (0.18-2.97)
If not quitting, help identify barriers	1.44 (0.60-3.41)
Advanced counseling	
Helping smokers prepare to quit[b]	1.0 (0.82-1.21)
Helping smokers anticipate barriers[c]	1.07 (0.88-1.31)
Managing relapses[d]	0.88 (0.54-1.45)
Advising smokers to set a quit date	1.24 (0.66-2.35)
Getting social support[e]	1.04 (0.6-1.80)

[a]N/A: not applicable.

[b]Helping smokers prepare to quit codes include identify reasons and benefits for quitting; provide information for follow-up visits with doctor; and recommend over-the-counter nicotine patch, other medications.

[c]Help smokers anticipate barriers codes include anticipate challenges in the beginning; anticipate nicotine withdrawal; alcohol is associated with relapse; and review past quit attempts, what helped, and what led to relapse.

[d]Manage relapses codes include if relapse occurs, review and learn from experience; if relapse occurs, repeat quit attempts; if relapse occurs, reassess problems; and total abstinence is essential.

[e]Getting social support codes include get support from family, friends, and coworkers; and other smokers in house are not helpful.

Discussion

Principal Findings

Our discussion focuses on three topics: (1) approximately, one-third of the smokers who were offered an opportunity to asynchronously communicate with the TTS over the web chose to use the service, suggesting some level of interest in the service; (2) the coding analysis demonstrated that the TTSs were able to deliver best practice–based counseling on this asynchronous medium; and (3) the finding of no association between the amount of TTS use and smoking cessation has important implications.

Previous interventions with web-based counseling for a variety of behavioral and mental health conditions have had inconsistent results, with negative studies being a common finding. However, as these studies did not explore the message content, we cannot say whether these interventions delivered counseling that included evidence-based communication patterns. Thus, one goal was to understand the content of the messages. In our previous publication, we have qualitatively analyzed the communication between the smoker and the TTS and identified seven basic themes in the communication—talk about the process of quitting, barriers to quitting, reasons to quit, quit history, support and strategies for quitting, quitting with medication, and quit progress [15]. This paper adds to this evaluation by conducting a semiquantitative coding analysis and evaluating the association between the codes and the number of messages and smoking cessation. We also evaluated whether the content and volume of messages were related to the cessation outcomes at 6 months.

As noted, a third of our participants chose to message the TTS without actively prompting them. Please note that we did not require the smoker to message the TTS. In a study with smokers from Arkansas Children's Hospital, 21% of smokers participated following a fax referral [38]. Fax referral is a process in which smokers are referred to a quit line by simply faxing a referral form to the quit line. Quitline TTS, then, proactively calls the smoker to setup the counseling sessions. In an observational study of faxed referrals to the Ohio Tobacco Quit Line, 23% (n=1616) of smokers were able to be enrolled out of a total of 6951 faxed referrals [39]. These results may suggest that there is some merit to our assumption that asynchronous counseling offers a lower barrier to access, acknowledging that the participants in our study may be more motivated to quit than the general population of smokers, as indicated by their registration on WATI. Furthermore, among those who communicated with the TTS, the majority (32/45, 71%) indicated the communication was helpful.

We used the Motivational Interviewing Self-Evaluation Checklist and the SCC Scale for developing our codes to evaluate the TTS messaging. As noted above, the Motivational Interviewing Self-Evaluation Checklist was originally developed as a self-evaluation tool to improve motivational interviewing skills [28-32], while the SCC Scale was originally developed to help in assessing, improving, or testing evidence-based methodologies for SCC [33]. A contribution of our work is adapting these to code messages between TTSs and smokers

on the web. Some of the messages included motivational interviewing and other behavioral strategies. The techniques used included supporting client strengths and promoting engagement, among others. TTSs were able to include messaging to help smokers prepare to quit and anticipate barriers.

In our content analysis of the smokers' messages, we found that smokers' messages most often discussed motivations to quit and current and past treatments used for smoking cessation, including their previous use of nicotine replacement therapy and medications. This suggested that smokers were able to express their motivations to quit, as well as the challenges they face while quitting. In addition, smokers were also able to express what worked or did not work in their past treatment. This ability of the smokers to include these concepts in their messages is important as it allows the TTSs to tailor their messaging to the smokers. This was evident in our analysis of the TTS messaging, during which TTSs were able to incorporate behavioral strategies, offer advice on treatments, and highlight motivations to quit.

We did not find any association between smoking cessation and the codes in the TTS messages, the selected TTS, and the amount of TTS use. As noted above, only a third of our participants chose to message the TTS and even among these participants, the mean number of messages sent was 2.6. This may have contributed to our inability to detect an effect. The tobacco treatment guidelines recommend four or more sessions to increase the effectiveness of counseling [1]. A previous meta-analysis has also indicated that the session length is an important factor in the effectiveness of counseling [1]. For the light-touch, low-intensity approach of asynchronous counseling, the number of interactions may need to be higher. In future work, we plan to explore approaches to increase engagement, both in terms of overall use of asynchronous counseling and the number of interactions between TTSs and smokers. First, accessing the TTS by logging into the web interface may be a barrier. Future studies could test whether text messaging and mobile phone–based access further reduce barriers to access. Second, users may need tailored and proactive messaging to continually motivate and encourage them to participate in asynchronous counseling. As noted, our study identified certain characteristics that were significantly associated with participation; compared with those who did not message the TTS, smokers who messaged the TTS were those who allowed smoking at home and who had visited a WATI before this study. These characteristics could potentially be used to tailor messages to motivate the smokers to engage in asynchronous counseling with the TTS.

Future trials should test asynchronous counseling in combination with synchronous counseling, as asynchronous counseling may be most beneficial when used as an augmentation to synchronous counseling. For example, the asynchronous counseling could be used in between synchronous counseling sessions to continually support the smoker. The timing (whether the synchronous counseling session precedes or follows the start of the asynchronous counseling) and the number of synchronous counseling sessions need to be tested. Furthermore, it is difficult to find an overall impact of TTS use because the content of the messages varied considerably.

Limitations

This was an observational study. Furthermore, not all smokers who participated in the larger randomized clinical trials participated in this study. Although this circumstance helped identify those who are likely to use the service, this may have caused some selection bias. The number of interactions between the TTSs and the smokers were also low. As this was a secondary, exploratory analysis, our study had limited power. Comparing TTS use=0 with TTS use=top quartile, we found a difference in cessation of 3% (top quartile: 9/56, 16%; 0 use: 62/480, 12.9%). With the sample available, we were powered to detect a difference of 11%.

Conclusions

Asynchronous counseling may be able to augment synchronous counseling, providing smokers the ability to connect with counselors at their convenience. We did not find an association between asynchronous counseling and smoking cessation. There are several potential explanations for this finding, including a lack of intervention engagement by participants, a lack of intervention fidelity (the counselors did not deliver evidence-based support), and a lack of power. We have now demonstrated that evidence-based cessation counseling did occur. Our discussion outlines several approaches to potentially increasing the effectiveness of asynchronous counseling, including a discussion of approaches to increase participant engagement, and testing the asynchronous counseling in combination with synchronous counseling. Future studies are needed to address the best ways to use asynchronous counseling for helping smokers.

Acknowledgments

Funding for this study was received from the National Institutes of Health grant nos. R01-CA-12909, U01-DE-16746, U01-DE-16747, U19-DE-22516, and U19-DE-28717. TH is also supported by the VA eHealth Quality Enhancement Research Initiative that he directs. RS is funded by a National Cancer Institute Career Development Award (K07-CA-172677). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the US government.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Decide2Quit website.
[DOC File , 1413 KB - jmir_v22i5e13289_app1.doc]

Multimedia Appendix 2
Topical content coding schema.
[DOC File , 150 KB - jmir_v22i5e13289_app2.doc]

Multimedia Appendix 3
Email communications between smokers and TTS.
[DOC File , 78 KB - jmir_v22i5e13289_app3.doc]

References

1. Fiore MC, Jaén CR, Baker TB, Bailey WC, Benowitz NL, Curry SJ, et al. The Tobacco Dependence Program (TDP). Rockville, MD: US Department of Health and Human Services. Public Health Service; 2008 May. Clinical Practice Guideline. Treating Tobacco Use and Dependence: 2008 Update URL: http://www.tobaccoprogram.org/clientuploads/documents/Consumer%20Materials/Clinicians%20Systems%20Mat/2008-Guidelines.pdf [accessed 2020-02-12]
2. Centers for Disease Control and Prevention. 2014 Nov 28. Current Cigarette Smoking Among Adults — United States, 2005–2013 URL: https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6347a4.htm [accessed 2017-12-05]
3. Centers for Disease Control and Prevention. Quitting Smoking URL: http://www.cdc.gov/tobacco/data_statistics/fact_sheets/cessation/quitting/index.htm [accessed 2017-12-05]
4. Fiore MC. Treating tobacco use and dependence: an introduction to the US Public Health Service Clinical Practice Guideline. Respir Care 2000 Oct;45(10):1196-1199. [Medline: 11203101]
5. Sheffer C, Brackman S, Lercara C, Cottoms N, Olson M, Panissidi L, et al. When Free Is Not for Me: confronting the barriers to use of free quitline telephone counseling for tobacco dependence. Int J Environ Res Public Health 2015 Dec 22;13(1):ijerph13010015 [FREE Full text] [doi: 10.3390/ijerph13010015] [Medline: 26703662]
6. Johnston LD, Miech RA, O'Malley PM, Bachman JG, Schulenberg JE, Patrick ME. ERIC - Education Resources Information Center. Monitoring the Future National Survey Results on Drug Use, 1975-2017: Overview, Key Findings on Adolescent Drug Use URL: https://eric.ed.gov/?id=ED589762 [accessed 2019-07-10]
7. Pew Research Center. 2019 Jun 12. Internet/Broadband Fact Sheet: Internet Use Over Time URL: http://www.pewinternet.org/data-trend/internet-use/internet-use-over-time/ [accessed 2019-08-10]
8. Pew Research Center.: Pew Research Center; 2015 Jun 10. Broadband vs Dial-up Adoption Over Time URL: https://www.pewresearch.org/internet/chart/broadband-vs-dial-up-adoption-over-time/ [accessed 2019-08-12]
9. Smith A. Pew Research Center. 2015 Apr 1. US Smartphone Use in 2015 URL: http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/ [accessed 2019-06-12]
10. Friederichs SA, Oenema A, Bolman C, Guyaux J, van Keulen HM, Lechner L. Motivational interviewing in a web-based physical activity intervention: questions and reflections. Health Promot Int 2015 Sep;30(3):803-815. [doi: 10.1093/heapro/dat069] [Medline: 24101160]
11. Ramo DE, Thrul J, Chavez K, Delucchi KL, Prochaska JJ. Feasibility and quit rates of the tobacco status project: a Facebook smoking cessation intervention for young adults. J Med Internet Res 2015 Dec 31;17(12):e291 [FREE Full text] [doi: 10.2196/jmir.5209] [Medline: 26721211]
12. Schaub MP, Wenger A, Berg O, Beck T, Stark L, Buehler E, et al. A web-based self-help intervention with and without chat counseling to reduce cannabis use in problematic cannabis users: three-arm randomized controlled trial. J Med Internet Res 2015 Oct 13;17(10):e232 [FREE Full text] [doi: 10.2196/jmir.4860] [Medline: 26462848]
13. McKay HG, Danaher BG, Seeley JR, Lichtenstein E, Gau JM. Comparing two web-based smoking cessation programs: randomized controlled trial. J Med Internet Res 2008 Nov 18;10(5):e40 [FREE Full text] [doi: 10.2196/jmir.993] [Medline: 19017582]
14. Patrick H, Fujii CA, Glaser DB, Utley DS, Marler JD. A comprehensive digital program for smoking cessation: assessing feasibility in a single-group cohort study. JMIR Mhealth Uhealth 2018 Dec 18;6(12):e11708 [FREE Full text] [doi: 10.2196/11708] [Medline: 30563807]
15. Williams JH, DeLaughter K, Volkman JE, Sadasivam RS, Ray MN, Gilbert GH, QUIT-PRIMONational Dental PBRN Collaborative Groups. Exploring Online Asynchronous Counseling With Tobacco Treatment Specialists in the QUIT-PRIMO and National Dental PBRN HI-QUIT Studies: Who Uses It and What Do They Say? Am J Health Promot 2018 Jun;32(5):1170-1177 [FREE Full text] [doi: 10.1177/0890117116670972] [Medline: 29848011]
16. Houston TK, Sadasivam RS, Ford DE, Richman J, Ray MN, Allison JJ. The QUIT-PRIMO provider-patient Internet-delivered smoking cessation referral intervention: a cluster-randomized comparative effectiveness trial: study protocol. Implement Sci 2010 Nov 17;5:87 [FREE Full text] [doi: 10.1186/1748-5908-5-87] [Medline: 21080972]

17. Houston TK, Sadasivam RS, Allison JJ, Ash AS, Ray MN, English TM, et al. Evaluating the QUIT-PRIMO clinical practice ePortal to increase smoker engagement with online cessation interventions: a national hybrid type 2 implementation study. Implement Sci 2015 Nov 2;10:154 [FREE Full text] [doi: 10.1186/s13012-015-0336-8] [Medline: 26525410]
18. Ray MN, Funkhouser E, Williams JH, Sadasivam RS, Gilbert GH, Coley HL, National Dental PBRN Collaborative Group. Smoking-cessation e-referrals: a national dental practice-based research network randomized controlled trial. Am J Prev Med 2014 Feb;46(2):158-165 [FREE Full text] [doi: 10.1016/j.amepre.2013.10.018] [Medline: 24439349]
19. Sadasivam RS, Kinney RL, Delaughter K, Rao SR, Williams JH, Coley HL, National Dental PBRN Group, QUIT-PRIMO Collaborative Group. Who participates in web-assisted tobacco interventions? The QUIT-PRIMO and National Dental Practice-Based Research Network Hi-Quit studies. J Med Internet Res 2013 May 1;15(5):e77 [FREE Full text] [doi: 10.2196/jmir.2385] [Medline: 23635417]
20. Coley HL, Sadasivam RS, Williams JH, Volkman JE, Schoenberger Y, Kohler CL, National Dental PBRNQUITPRIMO Collaborative Group. Crowdsourced peer- versus expert-written smoking-cessation messages. Am J Prev Med 2013 Nov;45(5):543-550 [FREE Full text] [doi: 10.1016/j.amepre.2013.07.004] [Medline: 24139766]
21. Center for Tobacco Treatment Research & Training. Tobacco Training Programs at Our Center URL: http://www.umassmed.edu/tobacco/training/ [accessed 2020-02-12]
22. National Center for Education Statistics (NCES). Percentage of the population 3 to 34 years old enrolled in school, by age group: Selected years, 1940 through 2008 URL: http://nces.ed.gov/programs/digest/d09/tables/dt09_007.asp [accessed 2019-08-15]
23. DiFranza JR, Rigotti NA, McNeill AD, Ockene JK, Savageau JA, St Cyr D, et al. Initial symptoms of nicotine dependence in adolescents. Tob Control 2000 Sep;9(3):313-319 [FREE Full text] [doi: 10.1136/tc.9.3.313] [Medline: 10982576]
24. MAXQDA. URL: https://www.maxqda.com/ [accessed 2019-09-12]
25. Karim F, Johansson E, Diwan VK, Kulane A. Community perceptions of tuberculosis: A qualitative exploration from a gender perspective. Public Health 2011 Feb;125(2):84-89. [doi: 10.1016/j.puhe.2010.10.005] [Medline: 21288542]
26. Schmidt-Weitmann S, Bösch S, Weidt S, Brockes C. Psychological online consultation service attracts men: a new approach to close the gap between being ill and being treated. Telemed J E Health 2015 Jan;21(1):48-54. [doi: 10.1089/tmj.2013.0370] [Medline: 25459521]
27. Rollnick S, Miller WR. What is motivational interviewing? Behav Cogn Psychother 1995 Oct;23(4):325-334. [doi: 10.1017/s135246580001643x]
28. Possidente CJ, Bucci KK, McClain WJ. Motivational interviewing: a tool to improve medication adherence? Am J Health Syst Pharm 2005 Jun 15;62(12):1311-1314. [doi: 10.1093/ajhp/62.12.1311] [Medline: 15947131]
29. Brodie DA, Inoue A, Shaw DG. Motivational interviewing to change quality of life for people with chronic heart failure: a randomised controlled trial. Int J Nurs Stud 2008 Apr;45(4):489-500. [doi: 10.1016/j.ijnurstu.2006.11.009] [Medline: 17258218]
30. Erol S, Erdogan S. Application of a stage based motivational interviewing approach to adolescent smoking cessation: the Transtheoretical Model-based study. Patient Educ Couns 2008 Jul;72(1):42-48. [doi: 10.1016/j.pec.2008.01.011] [Medline: 18304775]
31. van Eijk-Hustings YJL, Daemen L, Schaper NC, Vrijhoef HJ. Implementation of Motivational Interviewing in a diabetes care management initiative in The Netherlands. Patient Educ Couns 2011 Jul;84(1):10-15. [doi: 10.1016/j.pec.2010.06.016] [Medline: 20638812]
32. TrainingTeams. Motivational Interviewing Self-Evaluation Checklist URL: http://www.trainingteams.org/Document/MI_SelfEvaluationChecklist.pdf [accessed 2020-02-10]
33. Newhouse RP, Himmelfarb CD, Liang Y. Psychometric testing of the smoking cessation counseling scale. J Nurs Scholarsh 2011 Dec;43(4):405-411. [doi: 10.1111/j.1547-5069.2011.01420.x] [Medline: 22018103]
34. Chan SS, Sarna L, Wong DC, Lam T. Nurses' tobacco-related knowledge, attitudes, and practice in four major cities in China. J Nurs Scholarsh 2007;39(1):46-53. [doi: 10.1111/j.1547-5069.2007.00142.x] [Medline: 17393965]
35. Hughes JR, Keely JP, Niaura RS, Ossip-Klein DJ, Richmond RL, Swan GE. Measures of abstinence in clinical trials: issues and recommendations. Nicotine Tob Res 2003 Feb;5(1):13-25. [Medline: 12745503]
36. Smolkowski K, Danaher BG, Seeley JR, Kosty DB, Severson HH. Modeling missing binary outcome data in a successful web-based smokeless tobacco cessation program. Addiction 2010 Jun;105(6):1005-1015 [FREE Full text] [doi: 10.1111/j.1360-0443.2009.02896.x] [Medline: 20148782]
37. Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. Stat Methods Med Res 2013 Jun;22(3):278-295. [doi: 10.1177/0962280210395740] [Medline: 21220355]
38. Sisterhen LL, Sheffer CE, Bursac Z, Fischer EP. Rates of enrollment in smoking cessation services following fax referrals from a children's hospital. Acad Pediatr 2010;10(3):200-204. [doi: 10.1016/j.acap.2010.03.001] [Medline: 20452567]
39. Willett JG, Hood NE, Burns EK, Swetlick JL, Wilson SM, Lang DA, et al. Clinical faxed referrals to a tobacco quitline: reach, enrollment, and participant characteristics. Am J Prev Med 2009 Apr;36(4):337-340. [doi: 10.1016/j.amepre.2008.12.004] [Medline: 19201150]

Abbreviations

IRR: interrater reliability
OR: odds ratio
SCC: smoking cessation counseling
TTS: tobacco treatment specialist
WATI: web-assisted tobacco intervention

Edited by G Eysenbach; submitted 04.01.19; peer-reviewed by R Paz Castro, J McClure, MS Aslam; comments to author 01.03.19; revised version received 24.05.19; accepted 28.01.20; published 06.05.20.

Please cite as:
Sadasivam RS, Kamberi A, DeLaughter K, Phillips B, Williams JH, Cutrona SL, Ray MN, Gilbert GH, Houston TK, QUITPRIMO, National Dental PBRN Collaborative Group
Secure Asynchronous Communication Between Smokers and Tobacco Treatment Specialists: Secondary Analysis of a Web-Assisted Tobacco Intervention in the QUIT-PRIMO and National Dental PBRN Networks
J Med Internet Res 2020;22(5):e13289
URL: https://www.jmir.org/2020/5/e13289
doi:10.2196/13289
PMID:32374266

Original Paper

Health Care Employees' Perceptions of the Use of Artificial Intelligence Applications: Survey Study

Rana Abdullah[1*], BSc; Bahjat Fakieh[1*], PhD

Information Systems Department, King Abdulaziz University, Jeddah, Saudi Arabia

[*]all authors contributed equally

Corresponding Author:
Bahjat Fakieh, PhD
Information Systems Department
King Abdulaziz University
Al-Solaimaniah District
Jeddah, 21589
Saudi Arabia
Phone: 966 126952000 ext 67438
Email: bfakieh@kau.edu.sa

Abstract

Background: The advancement of health care information technology and the emergence of artificial intelligence has yielded tools to improve the quality of various health care processes. Few studies have investigated employee perceptions of artificial intelligence implementation in Saudi Arabia and the Arabian world. In addition, limited studies investigated the effect of employee knowledge and job title on the perception of artificial intelligence implementation in the workplace.

Objective: The aim of this study was to explore health care employee perceptions and attitudes toward the implementation of artificial intelligence technologies in health care institutions in Saudi Arabia.

Methods: An online questionnaire was published, and responses were collected from 250 employees, including doctors, nurses, and technicians at 4 of the largest hospitals in Riyadh, Saudi Arabia.

Results: The results of this study showed that 3.11 of 4 respondents feared artificial intelligence would replace employees and had a general lack of knowledge regarding artificial intelligence. In addition, most respondents were unaware of the advantages and most common challenges to artificial intelligence applications in the health sector, indicating a need for training. The results also showed that technicians were the most frequently impacted by artificial intelligence applications due to the nature of their jobs, which do not require much direct human interaction.

Conclusions: The Saudi health care sector presents an advantageous market potential that should be attractive to researchers and developers of artificial intelligence solutions.

(J Med Internet Res 2020;22(5):e17620) doi:10.2196/17620

KEYWORDS

artificial intelligence; employees; healthcare sector; perception; Saudi Arabia

Introduction

Overview

Recently, health care systems in several countries have begun to rely on storage of patient information to provide the best quality of health care. Due to rapid technological developments, health care information technology solutions provide the capacity to store enormous volumes of patient data; however, appropriate utilization of this data is essential to enhance health care quality, improve decision making, and reduce costs [1,2]. Over the last decade, artificial intelligence (AI) has provided significant advancements in this regard [3].

Artificial intelligence technologies were developed to offer practical benefits in different areas including health care applications [4,5]. A common feature of AI is the replication of human intellectual functions. From the health care perspective, AI brings a "paradigm shift to health care, powered by increasing availability of health care data and rapid progress of analytics techniques" [6].

Despite its promise, health sector employees have mixed attitudes and feelings regarding the implementation of AI

technologies [7,8]. This study investigated the attitudes and perceptions of the emergence of AI technologies among health care employees in Saudi Arabia.

Related Studies

Sarwar et al [9] used a questionnaire distributed to 487 pathologist respondents from 54 countries to explore perspectives of AI implementation in clinical practice. Their findings revealed generally positive attitudes toward AI, with approximately 75% reporting excitement or interest in AI as a diagnostic tool for improving quality and efficiency in pathology workflows. About 80% of the participants predicted the introduction of AI technology into the pathology laboratory in the coming years [9].

According to Brougham and Haar [10], futurists expect that nearly one-third of all existing jobs could be replaced by AI, smart technology algorithms (STARA), and robotics by the year 2025. There is limited information, however, about how employees perceive these technological innovations within the scope of their own careers and how they are being prepared for these possible changes. STARA awareness was created as a new measure for this research, capturing the degree to which employees feel their job might be replaced by technology. As career progression and associated technology knowledge increase with age, age was also tested as a STARA moderator. By employing a mixed methods approach with 120 employees, STARA awareness was tested on a range of jobs and well-being. STARA awareness was inversely correlated with career satisfaction and organizational commitment, and directly correlated with depression, cynicism, and turnover intentions [10].

Alamanova [11] investigated the perceptions of AI among human resource (HR) professionals in the fields of leadership, consultancy, and recruitment. The results showed that HR professionals have different feelings about AI than they do about other new technologies, as they were excited about reducing manual workloads while remaining cautious about adding excessive functionality to computing machines. In addition to electronic human resource management applications, HR professionals had concerns regarding technology interfaces and pricing [11].

Zande [12] explored workforce perspectives on Robotic Process Automation (RPA) implementation by conducting 8 interviews with staff whose jobs are automated by RPA. The study concluded that employees perceived the implementation of RPA as positive before and after its implementation. Because of the simple nature of the automated processes, employees believed that RPA implantation reduced their workload. The employees also expected their jobs to become more diverse and interesting. They continued to feel positive after implementation while also expressing concerns about the occurrence of process errors and how to handle those errors when the availability of technology expertise is limited [12].

Maskara et al [13] investigated current and future applications as well as employee acceptance of AI in the medical field. The sample included 73 cardiologists, dentists, ophthalmologists, and surgeons. Phone-based and face-to-face interviews were conducted to understand respondents' perception and awareness of AI solutions. Most respondents were aware of AI interventions in use in their field and some were already harnessing AI themselves; however, while doctors were aware of the advantages of AI advantages, they also perceived disadvantages in the high cost and lack of human touch [14].

Oh et al [8] explored AI awareness among doctors and assessed their attitudes toward medical AI applications. An online questionnaire conducted and distributed to 669 participants showed that doctors have positive attitudes toward AI implementation in the medical field. Most of the surveyed physicians assumed that their roles will not be replaced by AI [8].

Van Ittersum et al [13] sought to understand technology acceptance through review study and qualitative model development. They found that several variables influence technology acceptance, including the technology itself, characteristics of the individual user, and features of the organization of technologies used in the work environment. Individual user characteristics and technology characteristics interact to influence acceptance in terms of attitudes, intentions, and behaviors. Other variables such as technophobia, knowledge, or prior experience can be changed through exposure or training and instruction. Companies therefore have an opportunity to influence acceptance. A similar logic applies to organizational user characteristics. Opportunities to influence acceptance are provided when the variables affecting that acceptance are understood, such as ease of use, complexity, and fun/enjoyment, which can also be influenced through marketing activities. Other factors such as privacy, risk, and compatibility can be considered during the design process to maximize acceptance by at least some user groups [13].

Although these studies concentrated on the employee perceptions of AI implementation in their fields, the influence of different job types was not considered, nor did these studies consider respondents' knowledge of AI, although such knowledge could significantly influence perception. Finally, there is a clear lack of research on this subject in the Arabian and Saudi context despite the revolution that this area is witnessing in this field. This study was intended to fill this gap and find answers to the following questions:

1. What is the level of employees' knowledge about AI in the health care sector?
2. What is the perception of AI implementation among employees in the health care sector in Saudi Arabia?
3. Does job type influences perceptions of AI implementation of?

Table 1 summarizes prior studies on the subject.

Table 1. Summary of related studies.

Author	Year	Country	Field	Study type	Sample size	Main results
Alamanova [11]	2018	Estonia	Human resources	Qualitative	5	• Varying feelings about AI[a] • Focus on technology price and interface
Zande [12]	2018	Netherlands	Jobs automated by RPA[b]	Qualitative	8	• Positive attitudes toward technology implementation • Participants showed little concern
Maskara et al [14]	2017	India	Medical field	Qualitative	73	• Most respondents aware of AI interventions • Some perceived disadvantages of AI
Oh et al [8]	2019	South Korea	Medical field	Quantitative	669	• Positive attitudes toward AI implementation in the medical field • Assumption that roles will not be replaced by AI

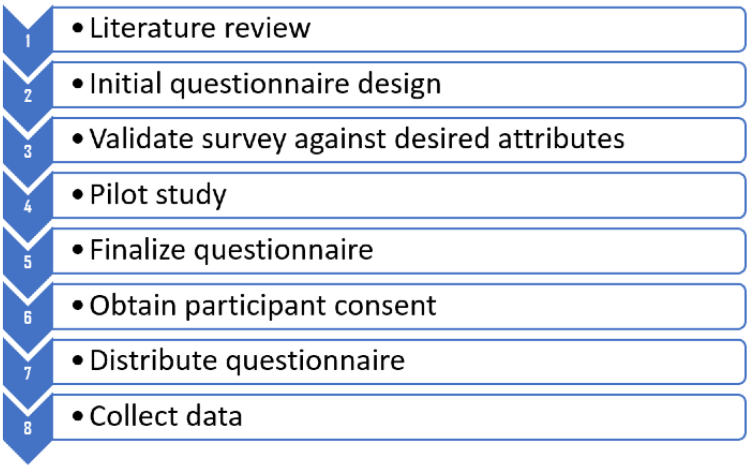
[a]AI: artificial intelligence.

[b]RPA: robotic process automation.

Methods

This study employed a descriptive analytical method and relied on a quantitative approach to collect primary data. The study was conducted from September 2018 to April 2019 at 4 of the main hospitals in Riyadh, Saudi Arabia. Figure 1 illustrates the study workflow. Each phase of the study will be discussed in detail in the following sections.

Figure 1. Survey workflow.



Population and Sample

The study population included all employees in public health care institutions in Saudi Arabia in the years 2018 and 2019, with a total of 300,699 working staff including physicians, nurses, pharmacists, and support staff [15]. The study sample consisted of 250 employees in the 4 largest hospitals in Riyadh, Saudi Arabia, which is home to 618 health institutions [16]. The sample included three types of employee: doctors, nurses, and technicians. Adequate sample size was met by convenience [17] and snowball sampling [18]. We employed Cohen's formula, which suggests that the effect size is low when the *r* value varies around 0.1, medium when the *r* value varies around 0.3, and large when the *r* value varies more than 0.5 [19]. The power analysis showed that a sample of 250 would provide an 80% chance of detecting correlations of ±0.223 at $P \leq .05$.

Questionnaire

The questionnaire was adapted from Oh et al [8] with revision, including changing the questions from multiple choice to a Likert scale and omitting questions related to pure medicine, as the original questionnaire targeted doctors. The questionnaire in its final form consisted of two main parts. The first part comprised a set of demographic questions to capture gender, age, job type, and level of educational attained. The second part of the questionnaire consisted of three sections. The first section (perceptions of AI) included four items: "I have a good knowledge about AI", "AI abilities are superior to the experience of humans", "AI could replace me in my job", "I have high hopes about AI applications in the health care sector". The second section (the advantages of using AI) included five items: "AI can speed up the process in health care"; "AI can help reduce medical errors"; "AI can deliver clinically relevant, vast amounts of high-quality data in real time"; "AI has no space-time constraint"; "AI suffers no emotional exhaustion or physical limitation". The final section (problems for AI application in health care) included five items: "AI cannot be used to provide opinions in unexpected situations", "AI is not flexible enough to be applied to every patient", "AI is difficult

to apply to controversial subjects", "AI has low ability to sympathize and consider the emotional well-being of the patient", "AI was developed by a specialist with little clinical experience in medical practice".

The questionnaire was validated by introducing it to a panel of experts in the medical and AI fields and circulated to a pilot sample of 15 employees from outside the sample population. Participants at this stage were asked to assess the clarity of the questions. Questionnaire reliability was established through the test-retest method, which is used to test changes when measuring a stable individual characteristic on different occasions [20]. The pilot study was resent to the same pilot sample of 15 employees after a 3-week period.

Data Analysis

Data were collected, categorized, and coded using Microsoft Excel (Microsoft Corporation) and then analyzed using Statistical Package for the Social Sciences (IBM Corporation). The mean, standard deviation, frequencies, and percentages were calculated for each question, while analysis of variance (ANOVA) was used to test for significant differences between the different demographic variables. Likert scores of 1-2.60 were considered low, 2.61-3.40 were considered moderate, and 3.41-5 were considered high.

Results

Table 2 shows the distribution of demographic variables among the questionnaire sample. Most of the sample (187/250, 74.8%) were female, as the majority of the sample were nurses (121/250, 48.4%) and between 20 and 40 years old (186/250, 74.8%). Moreover, nearly half of the respondents had a bachelor's degree (138/250, 55.2%).

Table 2. Respondent demographics (N=250).

Demographic and variable	Frequency, n (%)
Gender	
Male	63 (25.2)
Female	187 (74.8)
Age (years)	
20-30	132 (52.8)
31-40	54 (21.6)
41-50	58 (23.2)
>50	6 (2.4)
Job type	
Doctor	70 (28)
Nurse	121 (48.4)
Technician	59 (23.6)
Educational level	
Diploma	47 (18.8)
Bachelor	138 (55.2)
Postgraduate	65 (26)

Table 3 lists the respondents' answers regarding perception of AI technologies. The overall perception toward AI was moderate, with a mean of 3.01 (SD 1.13). More specifically, the item "AI could replace me in my job" ranked first with a mean of 3.11 (SD 1.17) or moderate on the Likert scale, while the least-applicable item was "I have good knowledge of AI" with a mean of 2.95 (SD 1.14) or moderate on the Likert scale.

Table 3. Perceptions of AI (N=250).

Rank	Question	Mean (SD)	*n*	Approximate agree rate	Level
4	I have good knowledge of AI[a]	2.95 (1.14)	185	74%	Moderate
2	AI abilities are superior to human experience	3.01 (1.17)	187	75%	Moderate
1	AI could replace me in my job	3.11 (1.13)	195	78%	Moderate
3	I have high hopes about AI applications in the health care sector	2.96 (1.11)	185	74%	Moderate
N/A[b]	Overall perception of AI	3.01 (1.13)	187	75%	Moderate

[a]AI: artificial intelligence.

[b]N/A: not applicable.

Table 4 lists respondents' answers regarding the advantages of using AI. Overall, the response was moderate with a mean of 3.36 (SD 1.16). The belief that "AI can speed up the process in health care" ranked first with a high level of acceptance (mean 3.50, SD 1.23). The lowest-ranked item was "AI can deliver clinically relevant, vast amounts of high-quality data in real time" with a mean of 3.24 (SD 1.17), a moderate level on the Likert scale.

Table 4. The advantages of using AI (N=250).

Rank	Question	Mean (SD)	*n*	Approximate agree rate	Level
1	AI can speed up the process in health care	3.50 (1.23)	175	70%	High
3	AI can help reduce the number of medical errors	3.36 (1.08)	167	67%	Moderate
5	AI can deliver clinically relevant, vast amounts of high-quality data in real time	3.24 (1.17)	162	65%	Moderate
2	AI has no space-time constraint	3.45 (1.17)	172	69%	High
4	AI has no emotional exhaustion or physical limitation	3.27 (1.16)	162	65%	Moderate
[a]N/A	The advantages of using AI overall perception	3.36 (1.16)	167	67%	Moderate

[a]N/A: not applicable.

Table 5 lists respondents' perceptions of problems in applying AI in health care. Generally, a moderate response was observed with a mean of 3.37 (SD 1.16). The response "AI is difficult to apply to controversial subjects" ranked first with a high mean of 3.62 (SD 1.17). The lowest-ranked item was a moderate response to "AI cannot be used to provide opinions in unexpected situations" with a mean of 3.20 (SD 1.14).

Table 5. The application of AI in health care (N=250).

Rank	Question	Mean (SD)	*n*	Approximate agree rate	Level
5	AI cannot be used to provide opinions in unexpected situations	3.20 (1.14)	160	64%	Moderate
4	AI is not flexible enough to be applied to every patient	3.28 (1.19)	165	66%	Moderate
1	AI is difficult to apply to controversial subjects	3.62 (1.17)	180	72%	High
3	AI has low ability to sympathize and consider the emotional well-being of the patient	3.34 (1.15)	167	67%	Moderate
2	AI was developed by a specialist with little clinical experience in medical practice	3.41 (1.17)	170	68%	High
[a]N/A	Problems regarding the application of AI in health care, overall perception	3.37 (1.16)	167	67%	Moderate

[a]N/A: not applicable.

Table 6 shows the results of the ANOVA test to find the statistically significant differences between the respondents' answers based on the demographic variables. There was no statistical difference by gender, age, or educational attainment; however, there were significant differences by job type (P=.007), with significance defined as .05. Overall, IT technicians in hospitals tended to have the most favorable opinions of AI, followed by doctors.

Table 6. ANOVA test of significance.

Variable	F value (*df*)	*P* value
Gender	0.942 (1, 248)	.341
Age	1.51 (3, 246)	.37
Job type	8.680 (2, 247)	.007[a]
Educational level	0.94 (2, 247)	.43

[a]Statistically significant difference.

Discussion

Principal Findings

The results of this study showed that employees in the Saudi health care sector have a moderate level of acceptance of AI applications, with most respondents indicating concern that their jobs would be replaced by AI. The results here are inconsistent with the study by Oh et al [8], who indicated that doctors do not believe they will be replaced by AI. The difference between these study findings may be attributed to the low level of knowledge claimed by the respondents in our study versus those in the prior study.

Speeding up health care processes was the main advantage identified by respondents in this study, consistent with prior studies that indicate AI can process a vast amount of data in an accurate, rapid, and efficient way by using complex statistical and computing algorithms [21,22]. In contrast, 163/250 (65%) of respondents did not believe AI can deliver clinically relevant, vast amounts of high-quality data in real time.

The most commonly perceived problem of AI in health care was difficulty in applying AI to controversial subjects and the least commonly identified problem was the inability of AI to provide opinions in unexpected situations, consistent with the findings of Oh et al [8]. The results of this study showed no significant differences in respondents' answers by gender, age, or educational level. There were significant differences by job type, as technicians were most likely to be exposed to the technological advances of AI, unlike nurses and doctors, who require direct human interaction with patients. This result is consistent with prior studies in which respondents expressed that in the future, computers and robots can do human jobs [23], although doctors feel they are not easily replaced [24].

This study had several limitations: only three job types in the health care sector were polled and the sample included only the largest hospitals in Riyadh. Other limitations include a limited sample size and the descriptive nature of the study. This study could be followed by additional studies including management-level job roles and smaller health care centers. Moreover, experimental studies may provide more realistic and comprehensive results.

Conclusion

This study assessed employee attitudes toward and perceptions of AI implementation in the health care sector. The study was conducted in 4 hospitals in Riyadh, the capital of Saudi Arabia. The results were mixed between fear of job replacement by AI and a lack of knowledge about AI technologies. Therefore, the results of this study indicate a need for training on the advantages, challenges, and issues surrounding the implementation of AI in health care and the potentials of these technologies to improve health care processes and efficiencies. Training would expand employees' knowledge of AI and their appreciation of its potential in the health care sector. Governments and universities can play significant roles in advancing health care toward utilizing AI technologies. In addition, the current status of AI use in health care in Saudi Arabia provides an attractive market for AI solution developers.

Conflicts of Interest

None declared.

References

1. Lovis C. Unlocking the Power of Artificial Intelligence and Big Data in Medicine. J Med Internet Res 2019 Nov 08;21(11):e16607 [FREE Full text] [doi: 10.2196/16607] [Medline: 31702565]
2. van Hartskamp M, Consoli S, Verhaegh W, Petkovic M, van de Stolpe A. Artificial Intelligence in Clinical Health Care Applications: Viewpoint. Interact J Med Res 2019 Apr 05;8(2):e12100. [doi: 10.2196/12100]
3. Yu K, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng 2018 Oct 10;2(10):719-731. [doi: 10.1038/s41551-018-0305-z]
4. Pannu A. Artificial Intelligence and its Application in Different Areas. International Journal of Engineering and Innovative Technology 2015 Apr;4(10):79-84.
5. Wu H, Chan N, Zhang CJP, Ming W. The Role of the Sharing Economy and Artificial Intelligence in Health Care: Opportunities and Challenges. J Med Internet Res 2019 Oct 15;21(10):e13469. [doi: 10.2196/13469]
6. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017 Jun 21;2(4):230-243. [doi: 10.1136/svn-2017-000101]

7. Fan W, Liu J, Zhu S, Pardalos PM. Investigating the impacting factors for the healthcare professionals to adopt artificial intelligence-based medical diagnosis support system (AIMDSS). Ann Oper Res 2018 Mar 19:1-26. [doi: 10.1007/s10479-018-2818-y]
8. Oh S, Kim JH, Choi S, Lee HJ, Hong J, Kwon SH. Physician Confidence in Artificial Intelligence: An Online Mobile Survey. J Med Internet Res 2019 Mar 25;21(3):e12422. [doi: 10.2196/12422]
9. Sarwar S, Dent A, Faust K, Richer M, Djuric U, Van Ommeren R, et al. Physician perspectives on integration of artificial intelligence into diagnostic pathology. npj Digit. Med 2019 Apr 26;2(1):28. [doi: 10.1038/s41746-019-0106-0]
10. Brougham D, Haar J. Smart Technology, Artificial Intelligence, Robotics, and Algorithms (STARA): Employees' perceptions of our future workplace. Journal of Management & Organization 2017 Jan 24;24(2):239-257. [doi: 10.1017/jmo.2016.55]
11. Alamanova M. The Perception of Artificial Intelligence and Other Technological Innovations Among Human Resources Specialists. Tallinn, Estonia: TALLINN UNIVERISTY OF TEHCNOLOGY; 2018.
12. Zande D. Robotic process automation: an employees? perspective: An exploratory case study on the perception towards Robotic Process Automation among employees. Amsterdam, Netherlands: Faculty of Economics and Business, University of Amsterdam; Jun 22, 2018.
13. ITTERSUM KOERT, ROGERS WENDY, CAPAR MUGE, CAINE KELLY, O'BRIEN MARITA, PARSONS LEONARD, et al. Understanding technology acceptance: Phase 1?literature review and qualitative model development. In: 2006, Georgia Institute of Technology. Atlanta, GA: Georgia Institute of Technology; Jan 2006.
14. Maskara R, Bhootra V, Thakkar D, Nishkalank N. A study on the perception of medical professionals towards artificial intelligence. International Journal of Multidisciplinary Research and Development 2017 Apr;4(4):34-39.
15. Al-Sulaiteen Z. Aleqtisadiah.: Saudi Research and Publishing Company; 2018 Jul 13. 41% of the workforce in the public health are foreign, that around 126 thousand URL: http://www.aleqt.com/2018/07/12/article_1419006.html [accessed 2019-11-12]
16. Ministry of Health. The Annual Report of Ministry of Health of 1439-1440 AH. Riyadh, Saudi Arabia: Saudi Ministry of Health; 2018.
17. Emerson RW. Convenience Sampling, Random Sampling, and Snowball Sampling: How Does Sampling Affect the Validity of Research? Journal of Visual Impairment & Blindness 2018 Dec 13;109(2):164-168. [doi: 10.1177/0145482x1510900215]
18. Woodley X, Lockard M. Womanism and Snowball Sampling:Engaging Marginalized Populations in Holistic Research. The Qualitative Report (TQR) 2016 Feb 15;21(2):321-329 [FREE Full text]
19. Cohen J. A power primer. Psychological Bulletin 1992;112(1):155-159. [doi: 10.1037/0033-2909.112.1.155]
20. Vilagut G. Test-Retest Reliability. In: Encyclopedia of quality of life and well-being research. Dordrecht: Springer; 2014:6622-6625.
21. Alsharqi M, Woodward WJ, Mumith JA, Markham DC, Upton R, Leeson P. Artificial intelligence and echocardiography. Echo Res Pract 2018 Dec 01;5(4):R115-R125 [FREE Full text] [doi: 10.1530/ERP-18-0056] [Medline: 30400053]
22. Shameer K, Johnson KW, Glicksberg BS, Dudley JT, Sengupta PP. Machine learning in cardiovascular medicine: are we there yet? Heart 2018 Jan 19;104(14):1156-1164. [doi: 10.1136/heartjnl-2017-311198]
23. Smith A, Anderson M. Washington: Pew Research Center. 2017 Oct 04. Automation in Everyday Life URL: https://www.pewresearch.org/internet/2017/10/04/automation-in-everyday-life/ [accessed 2019-11-26]
24. Krittanawong C. The rise of artificial intelligence and the uncertain future for physicians. European Journal of Internal Medicine 2018 Feb;48:e13-e14. [doi: 10.1016/j.ejim.2017.06.017]

Abbreviations

AI: artificial intelligence
ANOVA: analysis of variance
HR: human resources
RPA: robotic process automation
STARA: smart technology algorithms

Edited by G Eysenbach; submitted 27.12.19; peer-reviewed by M Alasnag, D Mendes; comments to author 25.02.20; revised version received 08.03.20; accepted 13.03.20; published 14.05.20.

Please cite as:
Abdullah R, Fakieh B
Health Care Employees' Perceptions of the Use of Artificial Intelligence Applications: Survey Study
J Med Internet Res 2020;22(5):e17620
URL: http://www.jmir.org/2020/5/e17620/
doi:10.2196/17620
PMID:32406857

Original Paper

An Exploration Into the Use of a Chatbot for Patients With Inflammatory Bowel Diseases: Retrospective Cohort Study

Aria Zand[1,2], MD; Arjun Sharma[1]; Zack Stokes[1], BS; Courtney Reynolds[1], MD, PhD; Alberto Montilla[3], PhD; Jenny Sauk[1], MD; Daniel Hommes[1,2], MD, PhD

[1]Vatche and Tamar Manoukian Division of Digestive Diseases, UCLA Center for Inflammatory Bowel Diseases, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA, United States

[2]Department of Digestive Diseases, Leiden University Medical Center, Leiden, Netherlands

[3]Cisco Systems Inc, Collaboration Technology Group, Dallas, TX, United States

Corresponding Author:
Aria Zand, MD
Vatche and Tamar Manoukian Division of Digestive Diseases
UCLA Center for Inflammatory Bowel Diseases
David Geffen School of Medicine, University of California at Los Angeles
10945 Le Conte Avenue
Los Angeles, CA, 90095
United States
Phone: 1 3102065403
Email: azand89@gmail.com

Abstract

Background: The emergence of chatbots in health care is fast approaching. Data on the feasibility of chatbots for chronic disease management are scarce.

Objective: This study aimed to explore the feasibility of utilizing natural language processing (NLP) for the categorization of electronic dialog data of patients with inflammatory bowel diseases (IBD) for use in the development of a chatbot.

Methods: Electronic dialog data collected between 2013 and 2018 from a care management platform (*UCLA eIBD*) at a tertiary referral center for IBD at the University of California, Los Angeles, were used. Part of the data was manually reviewed, and an algorithm for categorization was created. The algorithm categorized all relevant dialogs into a set number of categories using NLP. In addition, 3 independent physicians evaluated the appropriateness of the categorization.

Results: A total of 16,453 lines of dialog were collected and analyzed. We categorized 8324 messages from 424 patients into seven categories. As there was an overlap in these categories, their frequencies were measured independently as symptoms (2033/6193, 32.83%), medications (2397/6193, 38.70%), appointments (1518/6193, 24.51%), laboratory investigations (2106/6193, 34.01%), finance or insurance (447/6193, 7.22%), communications (2161/6193, 34.89%), procedures (617/6193, 9.96%), and miscellaneous (624/6193, 10.08%). Furthermore, in 95.0% (285/300) of cases, there were minor or no differences in categorization between the algorithm and the three independent physicians.

Conclusions: With increased adaptation of electronic health technologies, chatbots could have great potential in interacting with patients, collecting data, and increasing efficiency. Our categorization showcases the feasibility of using NLP in large amounts of electronic dialog for the development of a chatbot algorithm. Chatbots could allow for the monitoring of patients beyond consultations and potentially empower and educate patients and improve clinical outcomes.

(J Med Internet Res 2020;22(5):e15589) doi:10.2196/15589

KEYWORDS

chatbots; inflammatory bowel diseases; eHealth; artificial intelligence; telehealth; natural language processing

Introduction

Background

Recent technological advances have allowed for artificial intelligence (AI) to successfully integrate itself into many aspects of daily life. Besides implementation in voice bots such as Amazon's Alexa and Apple's Siri, AI is also utilized to predict financial stock market changes and answer student questions in educational settings [1]. In health care, AI is expected to disrupt the role of physicians as well; however, experts predict that AI will support the intelligence and knowledge base of physicians rather than replace them entirely [2]. For instance, AI can utilize deep-learning algorithms, which function like the neural networks of the brain and distinguish patterns, to recognize certain types of brain tumors, vascular conditions, or pneumonia on imaging scans and prioritize these cases in the workflow of a radiologist [2,3]. In addition, AI can be used to quickly review patient scans and rule out certain diagnoses, thereby increasing the efficiency and accuracy of a radiologist [2].

Another significant way AI can augment health care delivery is through medical chatbots. A chatbot, or chatterbot, attempts to simulate a natural conversation with a human user [4]. Medical chatbots are already being implemented into regular practice: the Insomnobot-3000 helps insomniacs get through the night, and the Endurance bot acts as a companion for dementia patients [5]. In addition, there are significant efforts toward the development of diagnostic chatbots. Some popular ones include Your.MD, Buoy Health, Sensely, Infermedica, and Florence (Table 1) [6].

Although there are limited data on these general medical chatbots in clinical practice, some independent bodies have provided preliminary and positive results in tests with more specific medical chatbots [7,8].

Most chatbots utilize natural language processing (NLP), which can be simply defined as the use of computers for analyzing human language [9]. One application of NLP relies on human identification of key elements within an event or situation that might constitute a useful summary of a given document or dataset [10]. Recently, there have been growing trends toward the use of electronic health records (EHRs). Multiple studies have attempted to use NLP to extract useful information from EHRs. In one study, researchers used NLP to identify patients with ulcerative colitis and Crohn disease from EHR data collected from Massachusetts General Hospital and Brigham and Women's Hospital [11]. The study developed an algorithm that partly relied on recognizing keywords associated with ulcerative colitis or Crohn disease to analyze the narrative texts and was verified via comparison to a physician's review and classification of the same narrative texts [11]. Ultimately, the study determined that NLP of patient narrative texts provided a more accurate means of identifying patients who had ulcerative colitis and Crohn disease than previous models that had relied on reviewing billing codes [11].

In another study by the University of Alabama, researchers developed an algorithm that analyzed the EHRs of patients collected over 3 years and organized the EHRs into pathology clusters based on key terms [12]. This team also concluded that electronic text mining of health records, or NLP, is an effective method for analyzing large health care datasets [12]. More recent studies have even attempted to use NLP models to study the semantics and sentence flows found in clinical narrative data [13,14]. The literature shows that it is common to perform exploratory analysis on natural language data to understand the topics and vocabulary of a specific domain in health care [9-14]. This exploration is often done by grouping keywords and categorizing topics or using open-source technology such as clinical Text Analysis and Knowledge Extraction [13]. A deep initial understanding facilitates the creation and comparison of more complex, health care-focused NLP models. However, it is worth noting that certain aspects of patient consultations in clinical settings, such as electronic record style, patient behavior, and physician experience, can vary from clinic to clinic [9,14]. This variability found within patient data puts limits on what NLP can do without a large and diverse sample.

In addition, despite the extensive literature on the topic, there seems to be a lack of research into the use of NLP to analyze raw consultation dialog data of patients with specific chronic conditions such as inflammatory bowel diseases (IBD). The organization of the patient with IBD to health care provider (HCP) dialog is likely to be distinct from a general patient population due to the complex nature of the disease. Understanding how these dialogs can be organized is an important first step in assessing the feasibility of a chatbot for this population.

Chatbots that utilize NLP can help to improve the way health care is delivered in multiple ways. For one, they improve accessibility to health care for patients outside of clinics and hospitals. From kids to the elderly, patients often need care outside of inpatient consultations; lack of such support is associated with inefficiency, high health care costs, and burdened HCPs [15]. With a chatbot, these patients would have immediate and autonomous support at home.

Table 1. Overview of current medical chatbots.

Name	Disease area	Objective	What does it do
Your.MD (UK[a])	General	Provide reliable information for common symptoms, recommends relevant resources	Safely advises patients based on symptoms described in an app-based messaging system
Endurance (Russia)	Dementia	Act as a companion for patients with short-term memory loss and help to identify signs of worsening patient condition	It works via voice recognition to ask questions and react to answers. It can speak on a variety of topics and pull interesting news from Google
Insomnobot-3000 (US[b])	Insomnia	Acts as a companion for insomniacs when they are awake at night.	Has conversations with patients via text
Pharmabot (Philippines)	Pediatrics	Designed to help pediatric patients get appropriate generic medicine for certain ailments	The system works in a software application that sets particular guidelines for interaction with the chatbot
Text-based healthcare chatbots on Mobile Coach (Switzerland)	Childhood obesity	Provide a peer character for obese teenagers and keep them engaged. In addition, sought to show the benefit of text-based chatbot interventions in health care	Works in a text channel within an app interface. Also, has predefined answer options for more efficient chat interactions
Molly by Sensely (US)	General	Diagnose patients with common ailments appropriately based on symptoms	Advises patients based on symptoms described in an app-based messaging system
Buoy Health (US)	General	Diagnose patients accurately based on symptoms. Harvard team developed the algorithm for this bot using 18,000 medical papers for data	Program asks a series of questions—for which there are predefined choices to choose from—to appropriately advise patient. Found on a Web-based software
Symptomate by Infermedica (Poland)	General	Attempt to increase health care provider efficiency, reduce costs, and improve patient flow by acting as a general symptom checker	Online software that collects and analyzes symptom data via predefined questions with answers to provide appropriate response
Florence (Germany)	General	Acts as a *personal nurse* that can remind patients to take prescriptions and keep track of user's health (weight, mood, etc)	Advises patients based on symptoms described in an app via Facebook messenger
Ada (international)	General	Help patients actively manage health based on common symptoms	Ada poses simple and relevant questions to patients and then compares their symptoms with thousands of similar cases to help provide possible explanations
Holly by Nimblr (US)	N/A[c]	Helps patients schedule and reschedule appointments to help prevent no shows or cancellations and improve patient experience	Interacts with patients via text and Amazon's Alexa to update electronic health records
Woebot (US)	Psychiatry	Make mental health care more accessible to people around the world	Uses methods from cognitive behavioral therapy to help patients think through situations. It also includes intelligent mood tracking

[a]UK: United Kingdom.

[b]US: United States.

[c]N/A: not applicable.

Objectives

The primary objective of this study was to accurately categorize large datasets of electronic messages between patients with IBD and HCPs using natural language processing (NLP) to assess the feasibility of developing a medical chatbot for patients with IBD.

Methods

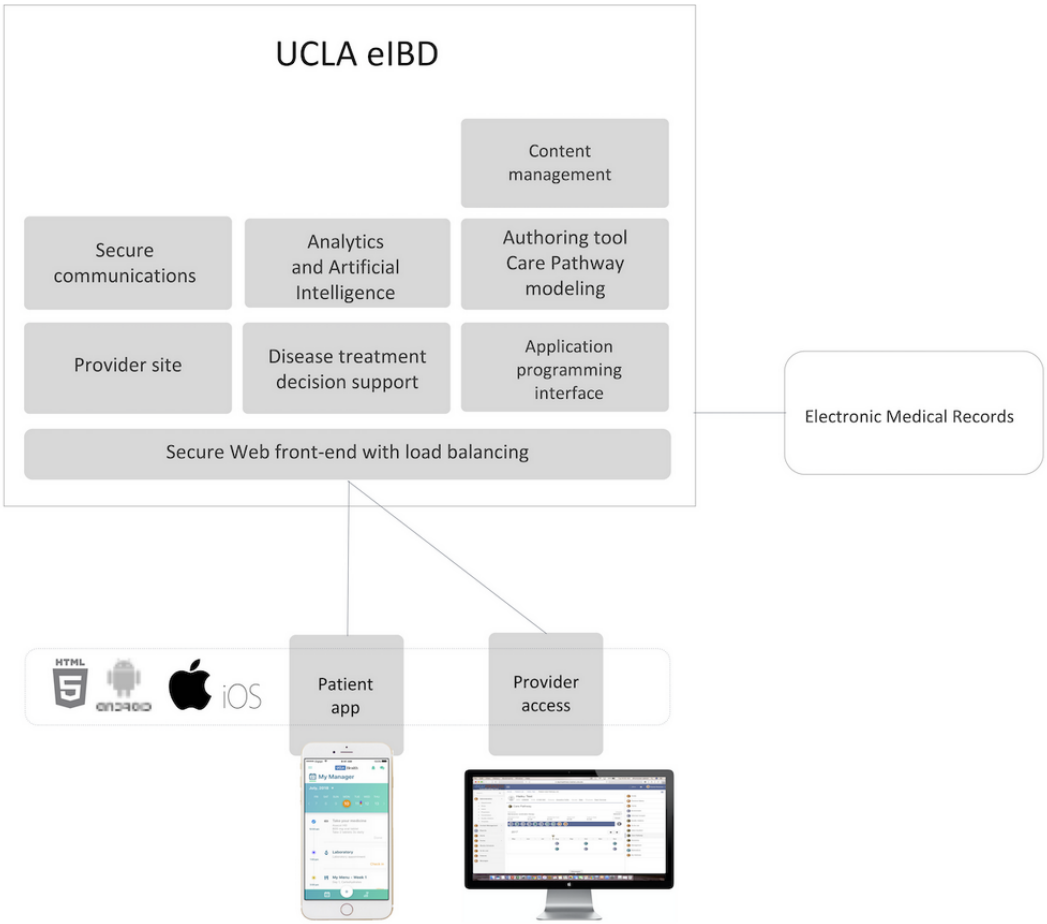
Design and Population

In this study, we aimed to assess the feasibility of utilizing NLP on historical electronic messaging data of patients with IBD for use in the development of a medical chatbot. As IBD is a chronic illness characterized by severe and recurring abdominal pain and diarrhea, patients require frequent contact with their physicians and care team to monitor these alternating disease states and potential relapses [16]. There is great potential here for a chatbot as patients need frequent monitoring beyond regular consultations, which is often troublesome due to the complex nature of the disease and a busy care team.

Patients enrolled in the University of California, Los Angeles (UCLA) Center for IBD electronic care management platform (UCLA eIBD) were retrospectively assessed. The UCLA eIBD platform is a care management software as a service with a Web-based platform for providers that includes treatment decision support, business intelligence, messaging functionality, and performance improvement tools. On the patient's side, there

is a mobile app that includes care management insight, educational modules, surveys, and messaging (Figure 1) [16]. Retrospective dialog data between patients and their care team from 2013 until 2018 was extracted and the feasibility of applying NLP categorization algorithms was assessed.

Figure 1. Overview of UCLA eIBD platform. AI: artificial intelligence; API: application programming interface.



All patients gave informed consent to participate. This study was approved by the Institutional Review Board (IRB) at UCLA with IRB protocol number 17-001208.

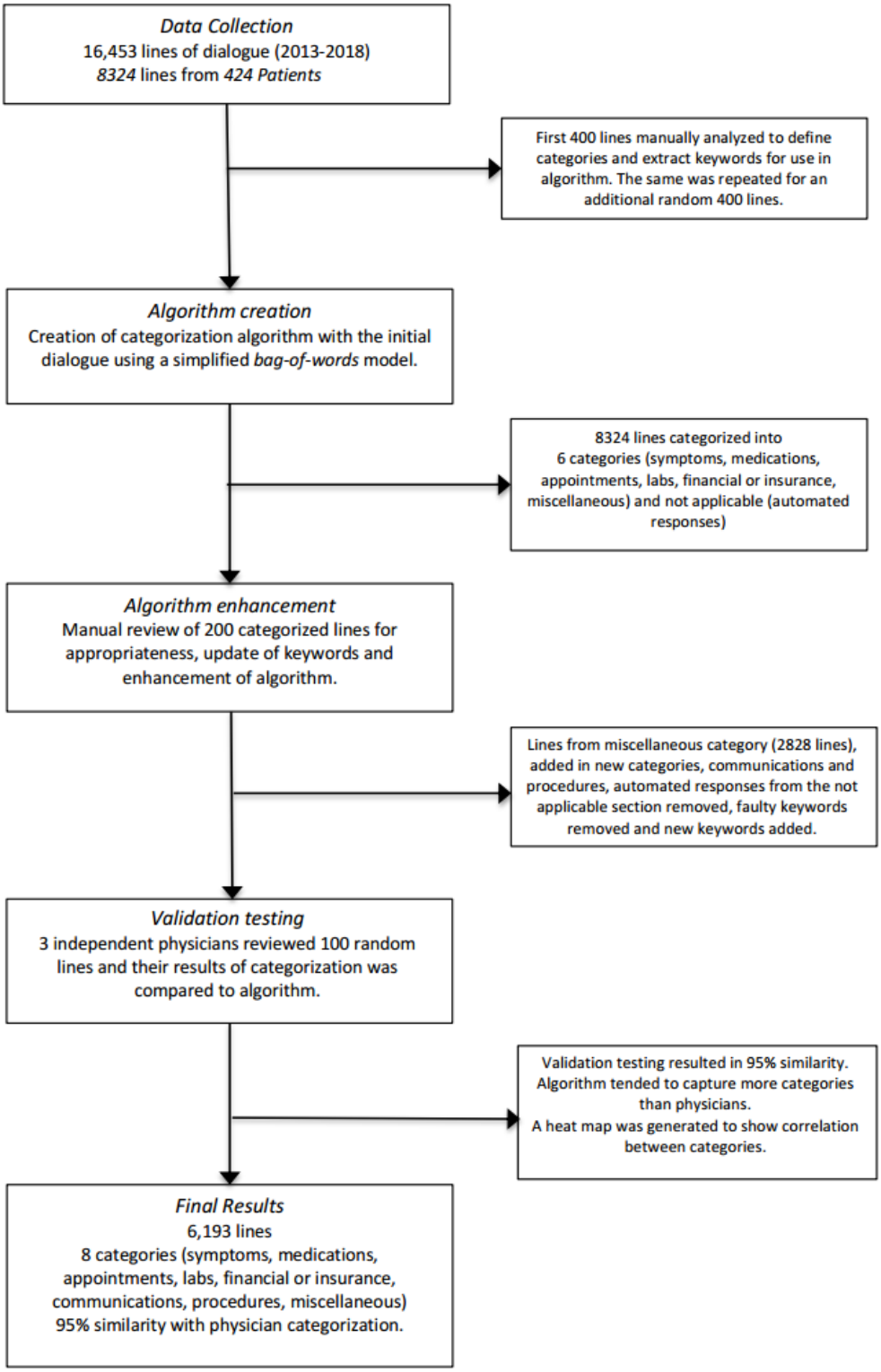
Data Collection and Anonymization

The dialogs were extracted from the UCLA eIBD database. The data consisted of the following: (1) a unique identifier, (2) first name, (3) last name, (4) date and time of message, (5) direction of message (HCP to patient or vice versa), (6) message content, (7) potential attachments, (8) HCP classification (urgent and nonurgent), (9) HCP action (responded yes or no), and (10) HCP response message content (Multimedia Appendix 1). The data were anonymized by removing the first and last names; for identification, we made use of the unique identifier in our analysis.

Categorization Method: Use of Natural Language Processing

Once the patient to HCP dialogs were stored in a Microsoft Excel sheet, the first 400 lines within the sheet were manually analyzed to identify relevant categories for use in our NLP algorithm. To clarify that the first 400 lines were representative, an additional 400 lines were randomly generated and manually reviewed as well (by AS and ZS). The analysis consisted of reading over each line to find an intent; if a particular intent was seen to occur frequently in these first lines, it was noted as a relevant category. The rationale behind using only categories observed in the sample was to make sure that the categories coded for were relevant to what the patient sample was discussing with their HCPs. Furthermore, 2 IBD gastroenterologists reviewed the categories found from the sample and reaffirmed that each category was representative of the IBD patient conversations they had encountered through electronic channels such as email. The same first 400 lines were then used to identify which keywords could assign a given dialog to a certain category (Multimedia Appendix 2). If a term appeared roughly 10 or more times in a given category, it was noted as a potential keyword; 2 physicians then reviewed and approved our list terms. Using these keywords, we employed a simplified, rule-based bag-of-words model to assign each line of dialog to the appropriate categories (Figure 2). The bag-of-words model essentially allows one to extract particular features of a text, that is, keywords, and score them with relevant numbers for modeling, or in our case, categorization [17]. To be certain, each line was converted into a standard bag-of-words with a score for each word in the form of a count of the number of times it appears within the line. With stop words removed, we extract the score of each keyword from all lines and assign to each line all categories for which any one keyword has a positive score.

Figure 2. Flowchart of inclusion and categorization. N/A: not applicable.



Enhancement and Correlation Assessment

On the basis of the preliminary results, the keywords of our initial categorization algorithm were refined, and new categories were created if necessary. If the categorization was not logical, we analyzed which keywords in the model miscategorized the dialog and made the necessary improvements. In addition, any uncategorized lines of dialogs were assigned a category, and their keywords were identified. The categorization algorithm was updated with the new, physician-approved keywords extracted from the uncategorized lines of dialog and the improvements of the existing categorization.

Once the code was refined to capture all the lines of dialog, a heat map was generated to showcase the overlap in categories, which refers to one line of dialog from a patient falling into two categories. It is worth noting that more than two categories could overlap, but there was no way to represent the higher levels of overlap in a relevant and concise diagram such as a

heat map. The goal was to paint a picture of what types of questions or concerns popped up together, which is instrumental in the actual development of a chatbot and creation of multicategory scenarios.

Validation of Accuracy

The accuracy of our categorization algorithm was tested by having 3 independent physicians from the UCLA Division of Digestive Diseases (AZ, CR, and DH) evaluate the appropriateness of the categorization. Each physician was assigned to categorize 100 randomly collected lines of dialog using the defined corresponding category number. In addition, the physicians categorized each line in the same style as the algorithm: numerical order with no spaces.

Once each of the doctors had finished categorizing the lines, the results were compared with the algorithm's categorization. We showcased the extent to which the algorithm and the doctors agreed or disagreed. To do this, the number of underclassifications and overclassifications the categorization algorithm made relative to the doctors' categories was calculated. For instance, if the algorithm missed a category that the doctor had, it would be counted as an underclassification of 1; if the category code had an extra category compared with the doctor, it would be counted as an overclassification of 1. We then created a bar chart plot based on this data. In addition, to understand the practicality of treating the doctors' assessments as ground truth, we computed the level of agreement between the three raters using Krippendorf alpha. This is a standard estimate of inter-rater reliability across ratings on a nominal scale.

To calculate a metric for the accuracy of the algorithm itself, we opted to use a nonstandard method of computing the success of the classification algorithm in an attempt to incorporate expert knowledge about the severity of misclassifications. As standard reliability measures such as Krippendorf alpha treat all disagreements between the raters and the algorithm with equal weight, we would not get a realistic view of the algorithm's strength across the spectrum of categories by following this approach. This was also done in an attempt to avoid aggregating our multiclass labels from the raters as doing so would put us at risk of destroying the variability in the ratings and inflating performance.

Software

Excel 2010 and R studio programming tool (R 3.4.0) were used for our analysis and algorithm creation (Multimedia Appendix 3).

Results

Data and Population Characteristics

Our sample consisted of 424 patients, 3 physicians, 3 nurses, and 2 administrative assistants with 16,453 lines of electronic dialog. Of the dialogs, 8324 lines were sent by 424 patients to their HCP (patient to HCP). Our analyzed patient cohort is 51.9% (220/424) female, 50.7% (215/424) have Crohn disease, and 46.9% (199/424) have ulcerative colitis with a mean disease duration of 13.4 (SD 10.4) years. The majority of the population is of the white (284/424, 67.0%) race and not of Hispanic or Latino ethnicity (386/424, 91.0%). Furthermore, most of the patients are employed (283/424, 66.7%) and have been enrolled in the care program for a mean of 4.6 (SD 1.3) years (Table 2).

Table 2. Characteristics of the inclusion cohort (N=424).

Variable	Values
Age (years), mean (SD)	42 (14)
Gender, n (%)	
Female	220 (51.9)
Male	204 (48.1)
Disease type, n (%)	
Crohn disease	215 (50.7)
Ulcerative colitis	199 (46.9)
Indeterminate colitis	10 (2.4)
Disease duration (years), mean (SD)	13.4 (10.4)
Race, n (%)	
White	284 (67.0)
Unknown	97 (22.9)
Asian	26 (6.1)
Black or African American	12 (2.8)
American Indian or Alaska Native	4 (0.9)
Native Hawaiian	1 (0.2)
Ethnicity, n (%)	
Not Hispanic or Latino	386 (91.0)
Hispanic or Latino	29 (6.8)
Unknown	9 (2.1)
Employment, n (%)	
Employed	283 (66.7)
Unemployed or unknown	141 (33.2)
Duration in program (years), mean (SD)	4.6 (1.3)

Algorithm Development and Initial Results

In our manual run-through of the first 400 out of the 8324 lines of dialog, we categorized them in six newly created and distinct categories: (1) medications, (2) symptoms, (3) appointments, (4) laboratory investigations, (5) finance/insurance, and (6) miscellaneous (lines that did not fall into any of the other categories). When the additional randomly generated 400 lines were reviewed for clarification, the same five relevant categories were found. At this point, we also kept a not applicable (N/A) section for automated responses produced by the mobile app itself that were in the dataset. For instance, “Patient has indicated there are no changes to medications.”

We identified what keywords were relevant to each of the categories (Multimedia Appendix 2). A categorization algorithm (bags-of-words model) was created based on the keywords extracted from the dialogs in the categories and applied to categorize the remaining lines of dialog.

Out of the 8324 lines of dialogs, the algorithm initially returned symptoms (1781/8324, 21.40% lines), medications (2114/8324, 25.40% lines), appointments (1781/8324, 21.40% lines), laboratory investigations (1648/8324, 19.80% lines), finance or insurance (358/8324, 4.30% lines), miscellaneous (2830/8324, 34.00% lines), and N/A (666/8324, 8.00% lines).

Enhancement of Natural Language Processing Categorization Algorithm

The miscellaneous section (2828/8317, 34.00% lines) was manually reviewed for 200 lines. The miscellaneous section was essentially randomly generated in that it was not organized by any dialog identifier, such as medical record number or patient name; it was simply the arbitrarily leftover dialogs from our initial run of the algorithm. As the dialogs here were short and not dominated by any one patient, we found it appropriate to review the first 200 lines as an accurate representation of the larger section. On review, two additional categories were identified within it: communications and procedures. In addition, the miscellaneous category was analyzed for keywords that would improve the scope of our initial categories. For instance, there were some medications we missed in our first test, such as Tylenol, that we were able to find upon review of the miscellaneous section and add as a keyword for medications. Furthermore, we removed keywords from the algorithm that were too general and inflated certain categories, such as the keyword *take* for the medications category.

Finally, the categorization algorithm was enhanced to remove dialog that only contained generic greetings, such as *Thank you* or *Hello*, and the automated responses from the N/A section from the dataset so that they did not affect the final counts. After this enhancement, 2131 lines were excluded and 6193 lines of dialog were left for categorization.

Final Natural Language Processing Categorization Results

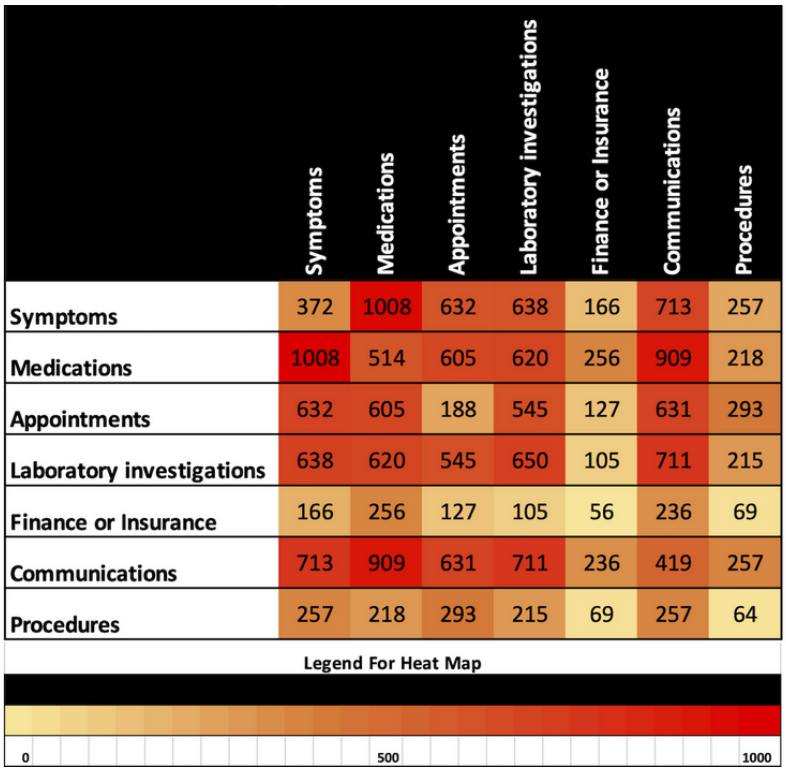
These refinements ultimately led to the algorithm yielding 32.83% (2033/6193) of the dialog relating to symptoms, 38.70% (2397/6193) to medications, 24.51% (1518/6193) to appointments, 34.01% (2106/6193) to laboratory investigations, 7.22% (447/6193) to finance or insurance, 34.89% (2161/6193) to communications, 9.96% (617/6193) to procedures, and 10.08% (624/6193) being miscellaneous (Table 3). The frequency of this overlap was measured for each possible pair combination of the categories and is displayed in a heat map (Figure 3). For instance, medications and symptoms appeared more together than they did on their own, as did communications and symptoms. Similarly, procedures and finance were very rarely brought up on their own (Figure 3).

Table 3. Final categorization results (N=6193).

Category	Percentage of total sample[a], n (%)
Symptoms	2033 (32.83)
Medications	2397 (38.70)
Appointments	1518 (24.51)
Laboratory investigations	2106 (34.01)
Finance or insurance	447 (7.22)
Communications	2161 (34.89)
Procedures	617 (9.96)
Miscellaneous	624 (10.08)

[a]These percentages represent how frequently these categories occur in the sample of dialogs. As the categories mostly overlap in the dialogs, the percentages do not add up to 100%.

Figure 3. Heat map of category overlaps in dialog. This map shows the frequency of category overlap in pairs and how often the categories occurred by themselves out of the 6193 dialogs. Note: across the diagonal, the map is a mirror of itself.



Validation of Natural Language Processing Accuracy

Three independent raters (AZ, DH, and CR) categorized 100 random lines of dialog, and their categorization was compared with our algorithms. The raters categorized in the exact style of the algorithm, so if the categories were symptoms, appointments, and medications, they would write *123*. Applying Krippendorf alpha to these assessment ratings, we get an

estimate of .61, indicating that there was moderate-to-high agreement between the doctors.

In our underclassification and overclassification representation of the chatbot's accuracy, we found that most of the errors were pooled at one difference, suggesting that the code and the doctors had a high level of agreement on most of the dialogs. Furthermore, the graph we constructed shows that the category code tended to over classify rather than under classify the subjects of the dialogs (Table 4). As one can see from the table, there is a significant drop in the instances of two or more underclassifications, with four to five missed categories having a frequency of 0 (Table 4). When we accounted for the 1 to 2 overclassifcation differences and the one category underclassification differences as minor, we found that 285 of the 300 tests had the program and physicians reasonably agreeing on categories. This meant that our code showed minor to no differences in 95% (285/300) of cases.

Table 4. Accuracy test results.

Number of categories added or missed by the algorithm in a given line	Instances in sample for overclassification, n	Instances in sample for underclassification
1	71	47
2	29	5
3	5	1
4	3	0
5	1	0

Discussion

Principal Findings

We were successful in categorizing large amounts of electronic messages between patients and providers into a reasonable number of categories (<10). Roughly 90.00% (5574/6193) of dialogs that came from patients fell into only seven categories, which shows potential for developing a chatbot with an NLP algorithm that can handle most IBD patient's questions and concerns. In addition, our heat map gave us insight into how these categories correlate with each other in the dialogs. In terms of chatbot development, this map allows a developer to be aware of what categories or topics tend to appear together in patient with IBD to HCP dialogs. This insight would allow the developer to better prepare the chatbot's NLP algorithm to identify topic transitions in a patient conversation and respond appropriately. In addition, our accuracy test supported the reliability of this result. Most of the differences recorded in our test (100/162, 61.0%) were simply due to code over classifying with one or two categories, but it rarely missed the primary intent (Table 4). Even when it did miss a category relative to the physician, the program was not necessarily incorrect upon review. For instance, one of the dialogs in the accuracy sample had a patient describing their symptoms or medications and subtly mentioning their laboratory investigations as their *previous averages*. Although the doctors recognized this and appropriately categorized the line as symptoms, medications, and laboratory investigations, the algorithm categorized it as symptoms and medications only, as averages was not a keyword we had programmed for laboratory investigations. Despite this, the program correctly identified the primary intent of the dialog, which is why we considered these types of differences minor in measuring the accuracy of our program.

Limitations

One limitation of this study is that our patient sample is fairly homogenous, consisting of mostly young (mean age 42 years) and white patients, which limits the generalizability of our results to other populations. In addition, most of the patients in the study are employed, which could have potentially changed the types of questions or concerns they expressed and the overall category distribution relative to other patient populations. It is also worth noting that we used the expert opinions of 2 IBD gastroenterologists to support the validity of the categories chosen and the selected keywords. This may affect the reproducibility of our results.

Comparisons With Prior Work

The next step from collecting data to developing a chatbot is to use machine learning methods to model the relationship between questions and responses [18]. Many chatbot knowledge bases (the database from which a chatbot draws its responses from) are hand constructed, which is time consuming and reduces the algorithm's versatility [19]. For instance, Artificial Linguistic Internet Computer Entity and *ELIZA*, two classic chatbots, utilize hand-constructed databases to generate a response that matches a given human input [20]. As an alternative, some developers have attempted to extract high-quality dialog data from online discussion forums to efficiently create a knowledge base for specific domain chatbots [19]. The purpose of collecting these dialog datasets is to give the chatbot a training ground to learn how to accurately respond to a specific domain of human input responses with minimal human fine tuning, or simply put: machine learning [18,21]. This machine learning approach also allows for the chatbot to continue learning through its interactions and improve its accuracy. Microsoft's Xiaoice chatbot has successfully applied this model and has already amassed a following of about 660 million online users [22]. When assessing the appropriateness of our data for actual chatbot development, our code could be distributed and tested in other centers with the same historical data without requiring much customization and would eliminate the need for hand-constructed databases.

Conclusions

Looking at the global trends of technology in health care, usage of smartphones and electronic health apps is on the rise [2,4,6].

Patient-provider communication through electronic messaging apps is becoming the standard. In our population, 25.0% (1518/6193) of messages were related to appointments. A chatbot could effectively automate requests regarding booking and cancellations or even play an instrumental part of triage, following the same guidelines as nurses, saving the provider team valuable time that could be redistributed to better patient care. The benefit is that a chatbot is available at all times, can handle tremendous amounts of conversation, and has no wait times.

Through the UCLA eIBD platform, we have already created a high-quality knowledge base of human dialogs that can be used to train an IBD chatbot using NLP. We showcased that it is feasible to categorize large amounts of electronic messaging data in one of the most complex chronic conditions into a reasonable number of categories. Given the feasibility of this categorization and the potential benefits of a chatbot, the next step would be to develop a chatbot and test it in a patient population with IBD. Further studies are required to showcase the effect on patients, providers, and costs and potential extrapolation to other chronic conditions.

Authors' Contributions

AZ was responsible for the study concept and design, analysis and interpretation of data, drafting of the manuscript, critical revision of the manuscript for important intellectual content, statistical analysis, and study supervision. AS was responsible for acquisition of data, drafting of the manuscript, statistical analysis, and critical revision of the manuscript for important intellectual content. ZS performed the statistical analysis and interpretation of the data. CR was responsible for the analysis and interpretation of data and the critical revision of the manuscript for important intellectual content. AM performed the revision of the manuscript for important intellectual content. JS was involved in the analysis and interpretation of data and the revision of the manuscript for important intellectual content. DH was in charge of the study concept and design, analysis and interpretation of the data, drafting of the manuscript, critical revision of the manuscript for important intellectual content, and the study supervision.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Section 1, Table of dialog data content.
[DOCX File , 12 KB - jmir_v22i5e15589_app1.docx]

Multimedia Appendix 2
Section 2, Table of keywords for categories.
[DOCX File , 14 KB - jmir_v22i5e15589_app2.docx]

Multimedia Appendix 3
Section 3, Copy of natural language processing algorithm code.
[DOCX File , 17 KB - jmir_v22i5e15589_app3.docx]

References

1. Phillips A. Becoming Human: Artificial Intelligence Magazine. 2018. 4 Industries Artificial Intelligence Is Transforming URL: https://becominghuman.ai/4-industries-artificial-intelligence-is-transforming-fe27b750769b [accessed 2018-10-17]
2. Kerschberg B. Forbes Magazine. 2018. How Real-Time AI is Accelerating the Disruption of Healthcare (Interview with Nuance Communications) URL: https://www.forbes.com/sites/benkerschberg/2018/03/19/__trashed-3/#15a28317530f [accessed 2018-10-17]
3. Brownlee J. Machine Learning Mastery. 2019. What is Deep Learning? URL: https://machinelearningmastery.com/what-is-deep-learning/ [accessed 2018-10-17]
4. Pearl R. Forbes Magazine. 2018. Artificial Intelligence In Healthcare: Separating Reality From Hype URL: https://www.forbes.com/sites/robertpearl/2018/03/13/artificial-intelligence-in-healthcare/#417555031d75 [accessed 2018-10-15]
5. Shewan D. WordStream: Online Advertising Made Easy. 2020. 10 of the Most Innovative Chatbots on the Web URL: https://www.wordstream.com/blog/ws/2017/10/04/chatbots [accessed 2018-11-15]
6. Sennaar K. Emerj-Artificial Intelligence Research and Insight. 2019. Chatbots for Healthcare–Comparing 5 Current Applications URL: https://emerj.com/ai-application-comparisons/chatbots-for-healthcare-comparison/ [accessed 2018-10-15]
7. Kowatsch T, Nißen M, Shih CH, Rüegger D, Volland D, Filler A, et al. Text-Based Healthcare Chatbots Supporting Patient and Health Professional Teams: Preliminary Results of a Randomized Controlled Trial on Childhood Obesity. In: 17th International Conference on Intelligent Virtual Agents. 2017 Presented at: IVA'17; August 27-30, 2017; Stockholm, Sweden.
8. Comendador BE, Francisco BM, Medenilla JS, Nacion SM, Serac TB. Pharmabot: a pediatric generic medicine consultant chatbot. J Automation Control Eng 2015;3(2):137-140. [doi: 10.12720/joace.3.2.137-140]

9. Iroju OG, Olaleke JO. A systematic review of natural language processing in healthcare. Int J Info Technol Comput Sci 2015 Jul 8;7(8):44-50. [doi: 10.5815/ijitcs.2015.08.07]
10. Turner AM, Liddy ED, Bradley J, Wheatley JA. Modeling public health interventions for improved access to the gray literature. J Med Libr Assoc 2005 Oct;93(4):487-494 [FREE Full text] [Medline: 16239945]
11. Ananthakrishnan AN, Cai T, Savova G, Cheng S, Chen P, Perez RG, et al. Improving case definition of Crohn's disease and ulcerative colitis in electronic medical records using natural language processing: a novel informatics approach. Inflamm Bowel Dis 2013 Jun;19(7):1411-1420 [FREE Full text] [doi: 10.1097/MIB.0b013e31828133fd] [Medline: 23567779]
12. Raja U, Mitchell T, Day T, Hardin JM. Text mining in healthcare. Applications and opportunities. J Healthc Inf Manag 2008;22(3):52-56. [Medline: 19267032]
13. Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, et al. Mayo clinical text analysis and knowledge extraction system (cTAKES): architecture, component evaluation and applications. J Am Med Inform Assoc 2010;17(5):507-513 [FREE Full text] [doi: 10.1136/jamia.2009.001560] [Medline: 20819853]
14. Kocaballi BA, Coiera E, Tong LH, White SJ, Quiroz JC, Rezazadegan F, et al. A network model of activities in primary care consultations. J Am Med Inform Assoc 2019 Oct 1;26(10):1074-1082 [FREE Full text] [doi: 10.1093/jamia/ocz046] [Medline: 31329875]
15. Petryszyn PW, Witczak I. Costs in inflammatory bowel diseases. Prz Gastroenterol 2016;11(1):6-13 [FREE Full text] [doi: 10.5114/pg.2016.57883] [Medline: 27110304]
16. van Deen WK, van der Meulen-de Jong AE, Parekh NK, Kane E, Zand A, DiNicola CA, et al. Development and validation of an inflammatory bowel diseases monitoring index for use with mobile health technologies. Clin Gastroenterol Hepatol 2016 Dec;14(12):1742-50.e7. [doi: 10.1016/j.cgh.2015.10.035] [Medline: 26598228]
17. Brownlee J. Machine Learning Mastery. 2017. A Gentle Introduction to the Bag-of-Words Model URL: https://machinelearningmastery.com/gentle-introduction-bag-words-model/ [accessed 2019-10-17]
18. Ramesh VC. Chatbots Magazine. 2016. Unsupervised Deep Learning for Vertical Conversational Chatbots URL: https://chatbotsmagazine.com/unsupervised-deep-learning-for-vertical-conversational-chatbots-c66f21b1e0f [accessed 2019-10-17]
19. Huang J, Zhou M, Yang D. Extracting Chatbot Knowledge From Online Discussion Forums. In: Proceedings of the 20th international joint conference on Artifical Intelligence. 2007 Presented at: IJCAI'07; January 6-12, 2007; Hyderabad, India p. 423-428. [doi: 10.5555/1625275.1625342]
20. Shum H, He X, Li D. From Eliza to XiaoIce: challenges and opportunities with social chatbots. Front Inf Technol Electron Eng 2018 Jan 8;19(1):10-26. [doi: 10.1631/fitee.1700826]
21. Yufeng G. Towards Data Science. 2017. The 7 Steps of Machine Learning URL: https://towardsdatascience.com/the-7-steps-of-machine-learning-2877d7e5548e [accessed 2019-10-17]
22. Zhou L, Gao J, Li D, Shum H. arXiv. 2019. The Design and Implementation of XiaoIce, an Empathetic Social Chatbot URL: https://arxiv.org/abs/1812.08989 [accessed 2019-10-17]

Abbreviations

AI: artificial intelligence
EHR: electronic health record
HCP: health care provider
IBD: inflammatory bowel diseases
IRB: Institutional Review Board
N/A: not applicable
NLP: natural language processing
UCLA: University of California, Los Angeles

Edited by G Eysenbach; submitted 22.07.19; peer-reviewed by A Kocaballi, I Gabashvili, J del Hoyo Francisco; comments to author 03.10.19; revised version received 27.11.19; accepted 01.03.20; published 26.05.20.

Please cite as:
Zand A, Sharma A, Stokes Z, Reynolds C, Montilla A, Sauk J, Hommes D
An Exploration Into the Use of a Chatbot for Patients With Inflammatory Bowel Diseases: Retrospective Cohort Study
J Med Internet Res 2020;22(5):e15589
URL: http://www.jmir.org/2020/5/e15589/
doi:10.2196/15589
PMID:32452808

Original Paper

Usability, Acceptability, and Effectiveness of Web-Based Conversational Agents to Facilitate Problem Solving in Older Adults: Controlled Study

Matthew Russell Bennion[1], BEng, MSc, PhD; Gillian E Hardy[1], BA, MSc, PhD; Roger K Moore[2], BA, MSc, PhD; Stephen Kellett[1], BSc, MSc, DClinPsy; Abigail Millings[1], BSc, PhD

[1]Department of Psychology, The University of Sheffield, Sheffield, United Kingdom

[2]Department of Computer Science, The University of Sheffield, Sheffield, United Kingdom

Corresponding Author:
Matthew Russell Bennion, BEng, MSc, PhD
Department of Psychology
The University of Sheffield
Cathedral Court
1 Vicar Lane
Sheffield,
United Kingdom
Phone: 44 07703049595
Email: m.bennion@sheffield.ac.uk

Abstract

Background: The usability and effectiveness of conversational agents (chatbots) that deliver psychological therapies is under-researched.

Objective: This study aimed to compare the system usability, acceptability, and effectiveness in older adults of 2 Web-based conversational agents that differ in theoretical orientation and approach.

Methods: In a randomized study, 112 older adults were allocated to 1 of the following 2 fully automated interventions: Manage Your Life Online (MYLO; ie, a chatbot that mimics a therapist using a method of levels approach) and ELIZA (a chatbot that mimics a therapist using a humanistic counseling approach). The primary outcome was problem distress and resolution, with secondary outcome measures of system usability and clinical outcome.

Results: MYLO participants spent significantly longer interacting with the conversational agent. Posthoc tests indicated that MYLO participants had significantly lower problem distress at follow-up. There were no differences between MYLO and ELIZA in terms of problem resolution. MYLO was rated as significantly more helpful and likely to be used again. System usability of both the conversational agents was associated with helpfulness of the agents and the willingness of the participants to reuse. Adherence was high. A total of 12% (7/59) of the MYLO group did not carry out their conversation with the chatbot.

Conclusions: Controlled studies of chatbots need to be conducted in clinical populations across different age groups. The potential integration of chatbots into psychological care in routine services is discussed.

(J Med Internet Res 2020;22(5):e16794) doi:10.2196/16794

KEYWORDS

transdiagnostic; method of levels; system usability; acceptability; effectiveness; mental health; conversational agents; older adults; chatbots; web-based;

Introduction

Background

The developers of psychological interventions have harnessed the internet as a delivery medium to enable increased access to evidence-based psychological therapies [1,2]. Psychological electronic therapies (e-therapies) have been defined and categorized in multiple ways that refer to properties, such as the type of technology being used or the level of therapeutic guidance involved [3]. E-therapies are typically grounded in cognitive behavioral therapy (CBT), as the protocol-driven format of CBT makes it a better fit for automation in comparison

with unstructured dynamic psychotherapies [4]. There is growing evidence indicating that e-therapies are clinically equivalent to traditional face-to-face therapies in reducing the symptoms of both common mental health problems and somatic disorders [5]. This evidence is based on the outcomes achieved with working-age adults. Therefore, this leaves older adults at risk of both digital and research exclusion. For example, although older participants are rarely excluded from clinical trials of e-therapies, they account for only 3% of participants [6]. Feasibility and pilot study evidence indicate that older adults are willing to use e-therapies [7] and do find the use of e-therapies a satisfying experience [8-10]. When tested, the evidence suggests that e-therapies can be clinically effective for older adults with symptoms of depression and anxiety [11-14].

An important consideration when designing e-therapies for older adults is the user experience of the technology. User experience research typically consists of assessments of the acceptability, usability, and satisfaction of the technology being used. User experience is defined as a "person's perceptions and responses resulting from the use and/or anticipated use of a product, system or service" [15] and usability as "the extent to which a product can be used by specified users to achieve specific goals with effectiveness, efficiency and satisfaction in a specified context of use" [15].

However, measuring the acceptability of e-therapies has typically been limited to only asking older adults to rate the acceptability of the technology before, during, and/or after using a program. Researchers have also assessed the user experience of e-therapies through measuring *treatment satisfaction*, but they have often used unvalidated questionnaires, thus bringing the results found into question [16].

Therefore, despite partially considering aspects of acceptability, usability, and satisfaction, it is rare for e-therapy studies to use the full array of international standards and associated validated instruments of usability, but there are some examples of good practice [17,18]. To maximize the reach and uptake of e-therapies for the older adults, adaptation of the methods for assessing user experience and system usability developed in engineering and computer science appears fit-for-purpose [19]. This is particularly important given the evidence that the older adults experience difficulty using e-therapies when instructions overload working memory, making it harder to effectively engage with the program [20]. Therefore, the older adults need to continually relearn how to use an e-therapy program, and on-going feelings of frustration would reduce the ratings of acceptability of the technology and risk disengagement [20].

Thus far, attempts to fully automate psychological therapies have been plagued with difficulties of low initial uptake and subsequent low adherence [21,22]. One method that has shown potential benefit for potentially increasing adherence to e-therapies is the use of conversational agents that deliver the content of e-therapies [23]. In this approach, software programs interpret and reply to lines of everyday normal language, and a therapeutic interaction is, therefore, created (ie, a conversation takes place between the client and *chatbot*, mirroring the conversation between the client and therapist). Therefore, the process of engaging with e-therapy is more personalized, dynamic, and bespoke, rather than simply following the psychoeducational exercises and self-monitoring that comprise most e-therapies.

In total, 2 conversational agents have subsequently been the focus of most research attention: ELIZA and Manage Your Life Online (MYLO), and these represent 2 differing theories and associated approaches to the treatment of emotional distress. The earliest attempt to develop a *chatbot* was by Joseph Weizenbaum in 1966. His program (*ELIZA*) was designed to mimic Rogerian counseling, a form of person-centered psychotherapy based on humanistic principles [24]. ELIZA applies simple natural language processing rules to the user's typed inputs to respond and generate text responses in the form of subsequent questions and responses appropriately. Despite its technical simplicity and the relative transparency of its therapeutic model, ELIZA can generate convincing dialogues, and there is anecdotal evidence of therapeutic effectiveness [25]. Despite the initial interest, little progress has been made to evolve and evaluate ELIZA into a fully automatic approach for treating mental health problems [4]. Another *chatbot* called MYLO has subsequently emerged. This is an attempt to implement a fully automated technique for treating mental health problems based on the principles of method of levels (MOL) therapy [26]. MOL is a transdiagnostic form of psychological therapy grounded in perceptual control theory [27]. MYLO uses open questions to encourage users to reflect on their thoughts, feelings, and behaviors, in a way that helps users to become more psychologically flexible, and thus, more adept at reducing distress [26]. MYLO simulates an MOL-style therapeutic conversation through an automated messaging interface.

There have been 2 previous trials with student populations comparing the outcomes achieved by MYLO and ELIZA from short single-session conversations. In a pilot trial (N=48) in a student population [28], MYLO was rated as more helpful and led to greater problem resolution, but there were no differences between the conversational agents with regard to any clinical outcomes (ie, depression, anxiety, and stress). In another student study (N=213), participants were randomized in a trial to either MYLO or ELIZA before completing poststudy and 2-week follow-up measures [29]. MYLO was again rated as significantly more helpful than ELIZA, but there were again similarly no differences between the conversational agents in terms of problem resolution and clinical outcomes.

To summarize, despite developments in the reliability of system usability testing in computer science and engineering, these approaches have not been consistently adopted in the context of the development and delivery of e-therapies. In addition, where e-therapies have been developed as conversational agents, any outcome evidence has also been unfortunately limited to working-age adults' samples. Therefore, more research is needed to investigate the clinical potential of conversational agents in the older adults.

Objectives

This study sought to compare and contrast the system usability of 2 *chatbots* (MYLO and ELIZA) in an older adult sample and to evaluate outcomes using a randomized and controlled

outcome methodology. We hypothesized that MYLO would be more acceptable, helpful, and usable than ELIZA, based on previous research [28,29], but there would be no difference in terms of clinical outcome. A secondary aim was to examine the relationship between the system usability and acceptability of the *chatbots*, particularly as Bird et al [29] specifically called for greater knowledge concerning the usability of MYLO in different groups.

Methods

Participants

Ethical approval was granted for the study (ref: 007599) by the University of Sheffield's Department of Psychology Ethics Committee. A study sample was recruited from the University of the Third Age (U3A), and participation was not monetarily incentivized. The U3A is a movement that aims to educationally stimulate members who have retired from work [30]. The study was advertised over the Web via U3A websites and offline via recruitment posters placed within U3A meeting places. Inclusion criteria for the study were (1) being older than 50 years, (2) being able to read and hear clearly (with glasses or hearing aids if necessary), (3) having no medically or professionally diagnosed current mental health disorder, and (4) currently experiencing a problem causing emotional distress.

Measures

The time points at which self-assessed measures were administered are summarized in a Standard Protocol Items: Recommendations for Interventional Trials diagram (Multimedia Appendix 1) and Table 1.

Participants provided a brief qualitative description of their *personal problems* and stated how long those problems had been occurring. Problem distress was measured on an 11-point Likert scale (from 0—not distressing at all to 10—highly distressing). Problem distress was measured at baseline, postintervention, and 2-week follow-up. Problem solvability was measured on an 11-point Likert scale (from 0—cannot be resolved to 10—easily resolved) at baseline. To measure problem resolution, participants rated on a Likert scale, at postintervention and 2-week follow-up, to what degree the problem had resolved (from 0—not resolved at all to 10—completely resolved).

Table 1. Summary and timeframe of measure administration.

Measure	Baseline	Postintervention	2-week follow-up
Problem distress	X[a]	X	X
Depression, anxiety, and stress scales 21	X	X	X
Problem solvability	X	—[b]	—
Problem resolution	—	X	X
Helpfulness	—	X	X
Use again	—	X	X
System usability scale	—	X	—

[a]The measure was taken at this time point.

[b]The measure was not taken at this time point.

Time

The time difference in minutes between the first and last timestamp of conversation logs was used to measure the duration of using the conversational agent.

Helpfulness

Participants rated how helpful the conversational agent was on an 11-point Likert scale (from 0—not helpful at all to 10—extremely helpful) at postintervention and at 2-week follow-up.

Use Again

Participants rated on an 11-point scale (from 0—most definitely not to 10—most definitely yes) the degree to which they would use the conversational agents again, but for a different problem, at postintervention and at 2-week follow-up.

The System Usability Scale

The system usability scale (SUS) measures perceptions of system technology and consists of a set of 10 statements scored on a 5-point scale [31]. An example item is "I found the system very cumbersome to use." SUS has been found to have high internal consistency in a number of large datasets [32,33], and it compares favorably with other usability measures [32]. An SUS score above 68 represents an above-average usability [34]. The SUS was only administered postintervention.

Depression, Anxiety, and Stress Scales 21

The depression, anxiety, and stress scales 21 (DASS-21) is a 21-item scale measuring depression, anxiety, and stress over the previous week on a 4-point scale [35]. Scores can range from 0 to 21 in each domain of the scale (depression, anxiety, stress) and are calculated by summing the scores of the representative 7 items. The DASS-21 has high internal consistency (depression: 0.91, anxiety: 0.84, and stress: 0.90[35]). Participants completed the DASS-21 at baseline, postintervention, and 2-week follow-up.

Procedure

To be involved, participants were required to either email or phone the lead researcher (MB). The researcher inputted each

potential participant's email address into a bespoke backend study management system; the system would then send participants emails containing a Web link to view the Web-based information sheet and consent form. Upon consenting, participants were sent a further email containing a set of instructions about each stage of the study, along with a Web link to allow them to begin interacting with the conversational agent (ie, participants were free to withdraw at this or any subsequent stage). Upon clicking the link, participants were taken to a set of self-assessment baseline measures within a Web-based questionnaire. After completion, the backend study management system randomly allocated, with equal probability, participants to either MYLO or ELIZA and generated the accompanying usernames, passwords, and program Web links to enable participants to access their allocated program.

The backend study management system would then email these details to the participants along with Web links to a user-guide video and usage tips Web page. The participants were given 24 hours in which they had to click the link in the email and log in to converse with their allocated conversational agent. Conversations were suggested to have a maximum duration of 20 min. After participants ended their conversation, the software presented a set of postintervention self-assessment measures within a Web-based questionnaire. Two weeks after completion, the backend study management system sent participants an email with a link to a Web-based questionnaire that contained the self-assessment follow-up measures.

Electronic Therapy Conversational Agents

To ensure that both systems were judged on the conversation they generated and not their respective user interfaces, the visual layout and input method of ELIZA were altered to mirror that of MYLO.

ELIZA

The implementation of ELIZA used in this study was based on a version by cyberpsych [36], which is accessible through the Web via a website hosted by the University of Sheffield. Conversations with ELIZA mimicked Rogerian client-centered counseling and aimed to facilitate problem solving by applying the core conditions for change during Rogerian counseling [24] (ie, congruence, empathy, and unconditional positive regard). ELIZA opens the session with *Hello, let's talk* and then adopts a consistent nondirective approach. The participants progress the conversation by typing their problems into a text input box and pressing the return key. ELIZA then responds with a question intended to maintain the conversation.

Manage Your Life Online

MYLO was accessed through the Web via a website hosted by the University of Sheffield. MYLO is an automated computer-based self-help program that mimics a therapeutic conversation between a client and a therapist using MOL as the change method. MYLO works by analyzing the participant's text input for key terms/themes and responds with questions aimed at encouraging conflict awareness and facilitating higher levels of awareness [28]. MYLO opens the session with *Please, tell me what's on your mind.* The participant progresses the conversation by typing their problem into a text input box and then clicking 1 of the response rating buttons. MYLO was developed by Warren Mansell at the University of Manchester.

Statistical Analysis

The study uses sample size calculations from Bird et al's study [29], which was a continuation of the work carried out by Gaffney et al [28]. A Cohen *d* of 0.79 was found for the baseline and postintervention comparison of distress scores of those in the MYLO group; a power analysis indicated that the minimum group size required was 19 with adequate power (0.8). Bird et al [29] found little differentiation in improvement in distress between groups (d=0.31). On the basis of this, the 2 conditions would, therefore, require a minimum sample size of 104. The study aimed to achieve the minimal power requirement, and a target to recruit 120 participants was set, which would result in 60 participants per group.

Data were analyzed using IBM SPSS for Microsoft Windows (version 24). The primary measure for the study was problem-related distress. DASS-21, problem resolution, time, use again, helpfulness, and system usability were secondary outcome measures.

The study used a mixed 2 × 3 analysis of variance (ANOVA), with the group (ELIZA or MYLO) as a between-participant factor and time (baseline, postintervention, and 2-week follow-up) as a within-participant variable for the primary outcome variable problem-related distress and secondary outcome measure DASS-21. Posthoc 2-tailed *t* tests were run to explore group differences using Bonferroni CI adjustment. Secondary outcome measures problem resolution, helpfulness, and use again were compared at postintervention and 2-week follow-up using ANOVA. Secondary outcome measures time and system usability were compared at postintervention using independent *t* tests that applied Bonferroni CI adjustment. To investigate the extent to which system usability was a predictor of problem resolution, helpfulness, and use again, a series of Pearson correlation coefficients were computed to assess the relationships between postintervention system usability, problem resolution, helpfulness, and use again. Simple linear regression was then carried out to determine the effect of postintervention system usability on postintervention helpfulness, use again, and problem resolution scores.

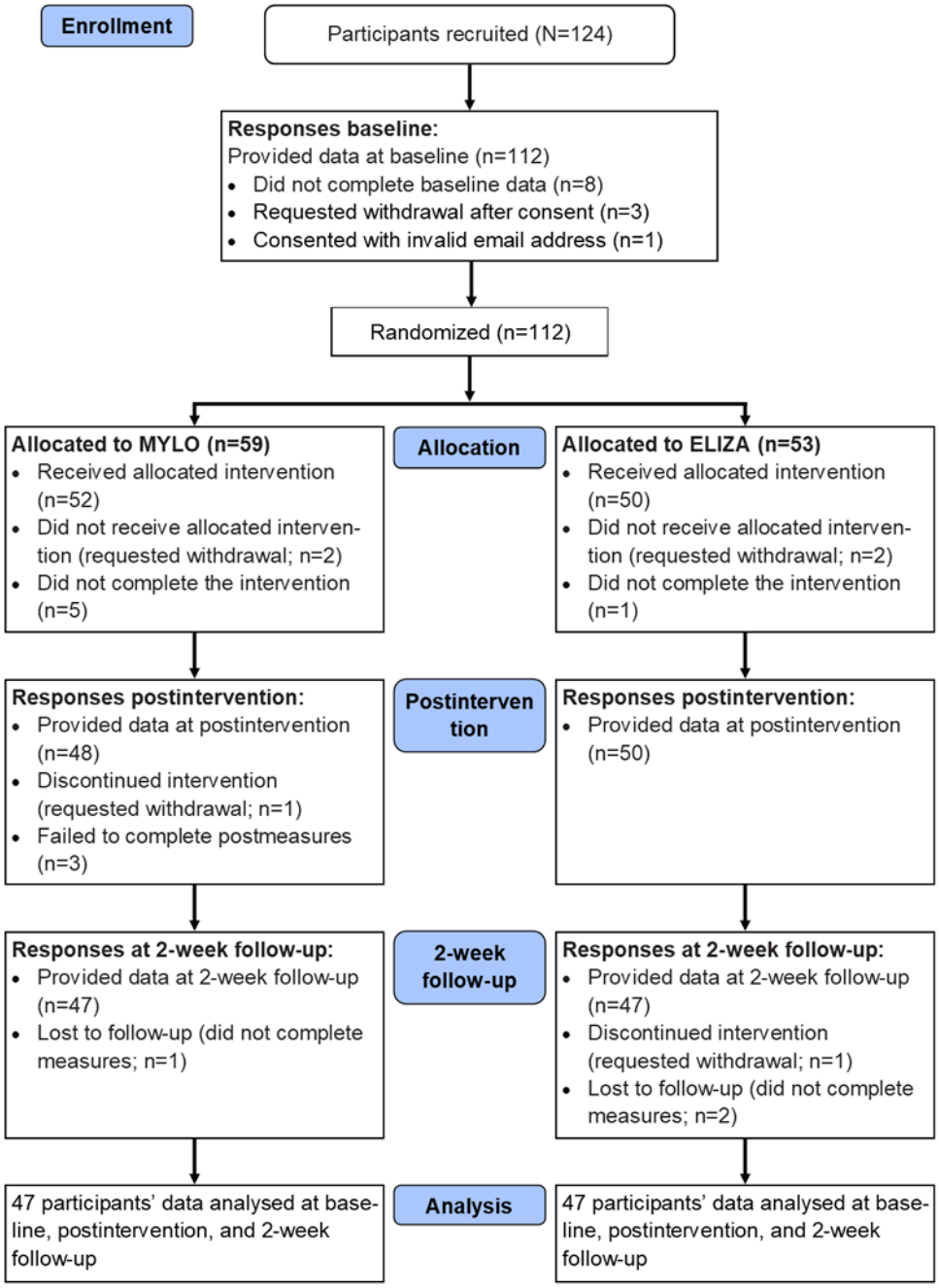
Results

Sample Characteristics

Age of the participants ranged from 51 to 90 years, with a mean of 69.21 (SD 6.76) years, and the study sample comprised 73.2% (82/112) females and 26.8% (30/112) males. A participant flow diagram is provided in Figure 1. In total, 112 participants completed baseline measures, were randomized, and then used the conversational agents, with 98 participants providing postconversation outcomes. Of the 59 participants allocated to MYLO, 52 completed the session with a dropout rate of 12% (7/59). Of the 53 participants allocated to ELIZA, 50 completed the session with a dropout rate 6% (3/53). Across both chatbots, 92.2% (94/102) participants completed the intervention. Of those who completed the intervention, 94 (MYLO: n=47 and ELIZA: n=47) provided outcomes across all 3 time points (ie,

baseline, postintervention, and 2-week follow-up). Those who completed the intervention had an average age of 68.4 (SD 6.49) years; 73% (69/94) of them were female and 27% (25/94) were male.

Figure 1. Participant flow diagram. MYLO: Manage Your Life Online.



Time Spent Using the Conversational Agents

The average amount of time spent engaged in conversation with MYLO was mean 24.17 min (SD 16.46), and the time spent in conversation engaged with ELIZA was mean 15.17 min (SD 8.77). On average, MYLO was used for 9 min longer than ELIZA (t_{92}=3.309; *P*<.001).

Problem Distress and Resolution

The problem-related distress and problem resolution scores for MYLO and ELIZA are reported in Table 2. There was no difference in reductions in problem-related distress over time between the 2 conversational agents ($F_{1,92}$=2.39; *P*=.13). There was a significant main effect of time on distress regardless of the conversational agent ($F_{2,84}$=55.85; *P*<.001). Problem distress significantly reduced between baseline and follow-up (*P*<.001), but there was no significant postintervention to follow-up reduction (*P*=.52). There was a significant interaction effect of the type of conversational agent and time on problem distress ($F_{2,84}$=3.21; *P*=.04), although this was a weak effect (eta-squared=0.03). This interaction was further investigated using *t* tests. The analysis showed that there was a significant

difference between interventions at follow-up (t_{92}=−2.013; *P*=.05), but no significant difference was found at baseline (t_{92}=0.428; *P*=.67) or postintervention (t_{92}=−1.593; *P*=.12). There were also no significant differences between the 2 conversational agents regarding their abilities to enable problem resolution ($F_{1,92}$=2.32; *P*=.13). There was a significant effect of time on problem resolution ($F_{1,92}$=15.87; *P*<.001).

Table 2. Mean (SD) for measures at baseline, postintervention, and 2-week follow-up.

Outcome measures	Manage Your Life Online (n=47), mean (SD)	ELIZA (n=47), mean (SD)
Problem distress		
Baseline	6.17 (1.55)	6.02 (1.81)
Postintervention	3.68 (2.14)	4.45 (2.51)
2-week follow-up	3.21 (2.23)	4.23 (2.67)
Problem solvability		
Baseline	4.09 (2.35)	3.55 (2.25)
Problem resolution		
Postintervention	2.17 (2.62)	1.51 (2.74)
2-week follow-up	3.77 (3.29)	3.04 (2.95)
Depression, anxiety, and stress scales 21 total		
Baseline	27.06 (16.18)	28.51 (19.17)
Postintervention	20.00 (14.59)	20.64 (15.04)
2-week follow-up	16.13 (13.91)	17.19 (14.71)
Helpfulness		
Postintervention	2.94 (2.89)	1.43 (1.86)
2-week follow-up	3.23 (2.81)	1.91 (2.21)
Use again		
Postintervention	4.21 (3.14)	2.45 (2.79)
2-week follow-up	4.43 (3.48)	2.70 (3.04)
System usability scale score		
Postintervention	63.56 (17.90)	56.97 (19.46)

Helpfulness, Use Again, and System Usability

There was a significant difference in helpfulness ratings over time between MYLO and ELIZA ($F_{1,92}$=8.801; *P*=.004). At postintervention, MYLO (mean 2.94, SD 2.89) was rated as significantly more helpful ($t_{78.661}$=3.016; *P*=.003) than ELIZA (mean 1.43, SD 1.86). There was a significant main effect of time on system helpfulness ratings ($F_{1,92}$=4.627; *P*=.03). In terms of use again ratings, there was a significant difference between the conversational agents ($F_{1,92}$=8.772; *P*=.004), with MYLO users postintervention more likely to use the conversational agent again for a future problem (t_{92}=2.882; *P*=.005). There was no main effect of time regarding the use again ratings ($F_{1,92}$=.816; *P*=.37). There were no significant differences in the postintervention system usability ratings between MYLO and ELIZA (t_{92}=1.710; *P*=.09). It is worth noting that the system usability scores for both MYLO (mean 63.56, SD 17.90) and ELIZA (mean 56.97, SD 19.46) were below the cut-off for an acceptable program (ie, <68).

Clinical Outcome

There was no statistically significant difference in DASS-21 scores over time between the conversational agents ($F_{1,92}$=0.139; *P*=.71). There was a significant main effect of time on total DASS-21 scores ($F_{1.830,168.368}$=33.538; *P*<.001). Total DASS-21 scores reduced significantly between baseline and postconversation (*P*<.001), between postconversation and follow-up (*P*=.02), and between baseline and follow-up (*P*<.001).

Usability and Acceptability of the Two Conversation Agents

There were statistically significant, moderate positive correlations between MYLO system usability ratings and postintervention ratings of helpfulness (r_{45}=0.546, *P*<.001) and interest in reusing MYLO (r_{45}=0.542, *P*<.001), and there was a statistically significant weak positive correlation between MYLO system usability ratings and problem resolution (r_{45}=0.420; *P*<.001; see Table 3 for details).

There was a statistically significant, weak positive correlation between the ELIZA system usability ratings and helpfulness

(r_{45}=0.344; P<.001) and interest in reusing ELIZA (r_{45}=0.387; P<.001) see Table 4 for details). Table 4 contains the helpfulness, use again, and SUS scores for MYLO and ELIZA.

There were statistically significant, moderate positive correlations between combined MYLO and ELIZA system usability ratings and postintervention ratings of the helpfulness of MYLO/ELIZA (r_{92}=0.473; P<.001) and interest in reusing MYLO/ELIZA (r_{92}=0.487; P<.001; see Table 5 for details).

Table 3. Pearson Correlations for postintervention Manage Your Life Online ratings of system usability, problem resolution, helpfulness, and willingness to use Manage Your Life Online again.

Variables	System usability scale score	Problem resolution	Helpfulness	Use again
System usability scale score	1	0.42[a]	0.55[a]	0.54[a]
Problem resolution	0.42[a]	1	0.78[a]	0.58[a]
Helpfulness	0.55[a]	0.78[a]	1	0.79[a]
Use again	0.54[a]	0.58[a]	0.79[a]	1

[a]Correlation is significant at the .01 level.

Table 4. Pearson Correlations for postintervention ELIZA ratings of system usability, problem resolution, helpfulness, and willingness to use ELIZA again.

Variables	System usability scale score	Problem resolution	Helpfulness	Use again
System usability scale score	1	0.11	0.34[a]	0.39[b]
Problem resolution	0.11	1	0.39[b]	0.26
Helpfulness	0.34[a]	0.39[b]	1	0.72[b]
Use again	0.39[b]	0.26	0.72[b]	1

[a]Correlation is significant at the .05 level.

[b]Correlation is significant at the .01 level.

Table 5. Pearson Correlations for postintervention Manage Your Life Online and ELIZA ratings of system usability, problem resolution, helpfulness, and willingness to use Manage Your Life Online/ELIZA again.

Variables	System usability scale score	Problem resolution	Helpfulness	Use again
System usability scale score	1	0.27[a]	0.47[a]	0.49[a]
Problem resolution	0.27[a]	1	0.61[a]	0.44[a]
Helpfulness	0.47[a]	0.61[a]	1	0.78[a]
Use again	0.49[a]	0.44[a]	0.78[a]	1

[a]Correlation is significant at the .01 level.

Further tests of MYLO using simple linear regression investigated the relationship between system usability score, helpfulness, use again, and problem resolution, with system usability scores as the predictor variable.

This revealed a significant relationship between the MYLO system usability score and helpfulness (P<.001). The slope coefficient for system usability was 0.088, so the resolution increased by 0.088 for each extra resolution point. The R^2=0.299 indicated that 29.9% of the variation in helpfulness was explained by the model containing only the system usability score a significant relationship between the MYLO system usability score and use again (P<.001). The slope coefficient for system usability was 0.095, so the resolution increased by 0.095 for each extra resolution point. The R^2=0.294 indicated that 29.4% of the variation in use again was explained by the model containing only the system usability score. There was also a significant relationship between the MYLO usability score and problem resolution (P=.003). The slope coefficient for system usability was 0.095, so the resolution increased by 0.095 for each extra resolution point. The R^2=0.176 indicated that 17.6% of the variation in problem resolution was explained by the model containing only the system usability score.

Tests of ELIZA using simple linear regression investigated the relationship between system usability score, helpfulness, use again, and problem resolution, with system usability scores as the predictor variable. This revealed a significant relationship between the ELIZA system usability score and helpfulness (P=.02). The slope coefficient for system usability was 0.033, so the resolution increased by 0.033 for each extra resolution point. The R^2=0.118 indicated that 11.8% of the variation in

helpfulness was explained by the model containing only the system usability score. There was also a significant relationship between the ELIZA system usability score and use again (P=.01). The slope coefficient for system usability was 0.055, so the resolution increased by 0.055 for each extra resolution point. The R^2=0.150 indicated that 15.0% of the variation in use again was explained by the model containing only the system usability score.

Finally, tests of MYLO and ELIZA results using simple linear regression investigated the relationship between system usability score, helpfulness, use again, and problem resolution, with system usability scores as the predictor variable. This revealed a significant relationship between system usability score and helpfulness (P<.001). The slope coefficient for system usability was 0.063, so the resolution increased by 0.063 for each extra resolution point. The R^2=0.224 indicated that 22.4% of the variation in helpfulness was explained by the model containing only the system usability score. A simple linear regression was used again to investigate the relationship between system usability score and use again, with system usability scores as the predictor variable. This revealed a significant relationship between system usability score and use again (P<.001). The slope coefficient for system usability was 0.080, so the resolution increased by 0.080 for each extra resolution point. The R^2=0.238 indicated that 23.8% of the variation in use again was explained by the model containing only the system usability score. There was also a significant relationship between usability score and problem resolution (P=.01). The slope coefficient for system usability was 0.038, so the resolution increased by 0.038 for each extra resolution point. The R^2=0.072 indicated that 7.2% of the variation in problem resolution was explained by the model containing only the system usability score.

Discussion

Principal Findings

The primary aim of this study was to compare the system usability, helpfulness, and effectiveness of 2 conversational agents (MYLO and ELIZA) with regard to problem solving within a nonclinical older adult sample. This study was, therefore, a replication and extension of previous studies [28,29], but this is the first study to compare these 2 conversational agents in an older adult sample. A secondary aim was to examine the relationship between system usability and acceptability of 2 differing *chatbots*. This is an important research because the ever-increasing demand for rapid access to psychological interventions in public services means that alternative delivery methods need to be considered and tested. Such methods can replace or supplement the traditional *high intensity-low throughput* approach of traditional one-to-one and face-to-face psychological therapy delivery. The conversational agents were grounded in differing theories and approaches to the resolution of psychological distress: MOL for MYLO [26] and humanistic counseling for ELIZA [24]. However, the conversational agents tended to enable problem resolution and reductions in problem-related distress, with MYLO showing significantly lower levels of problem-related distress at follow-up. In terms of clinical outcomes, each chatbot enabled immediate reductions in DASS-21, with reductions being improved over the follow-up period.

Participants spent significantly more time using MYLO, but it is worth noting that the time spent using the program was brief in either arm (ie, an average of 20 min and this was a prompt in the instructions for using the program). Average time spent using MYLO and ELIZA is just 10-min in working-age participants [29]. These results may indicate that adults aged above 50 years are more willing to try and converse with a program of this nature. The longer MYLO conversations may be a consequence of the program's more tailored and inquisitive questioning algorithm. In contrast, ELIZA has benefited from only limited improvements to its algorithm since its original implementation in 1966. The helpfulness and *use again* ratings of ELIZA and MYLO were significantly different, with MYLO being experienced as differentially more helpful and also more likely to be used again by participants. As MYLO was significantly more helpful, this may further explain why participants used MYLO for a significantly longer duration. These results mirror the evidence found in community working-age samples [28,29]. It may be the case that if time was allowed to be at the participant's discretion, then ELIZA may have been rated just as helpful as MYLO.

The second aim of this study was to investigate if system usability affected the acceptability of MYLO and ELIZA when used by the older adults. Generally, correlations between MYLO system usability and problem resolution, helpfulness, and interest in reusing the system were higher than those for ELIZA. These findings indicate that *chatbot* system usability has an impact on how users perceive and rate their experience of using a conversational agent. As Web-based delivery systems do not have the benefit of a therapist to explain the rationale for certain interventions, it is essential that system usability ratings are systematically collected over the developmental iterations of the systems. This is so that when a *chatbot* goes live, it is clear and easy to use. If a system is confusing or frustrating to use, then it is highly likely to be clinically ineffective; this arguably mirrors the evidence base concerning the therapeutic alliance in general psychotherapy [37].

The findings from this study appear consistent with accepted models of system usability (eg, International Organization for Standardization 2018 [38]). Although some previous studies have also used the SUS as a measure of system usability in e-therapies [39,40], it was a strength of this study to use this validated measure and is the first usage with an older adult population using a chatbot. It is worth noting that the theoretical underpinning of the 2 conversational agents (MOL versus humanistic counseling) may have influenced the perceptions of helpfulness and, therefore, the willingness to reuse the system. High rates of attrition are assumed to be a common problem with unsupported Web-based interventions, but a meta-analysis [41] has found that the percentage of completed sessions in face-to-face CBT (83.9%) did not differ from the percentage of completed sessions in internet-delivered CBT (80.8%). The overall session completion found in this study was higher 92.2% (94/102), but this was probably due to the intervention using a single-session approach.

Limitations and Future Directions

The study is limited by the fact that it did not recruit enough participants, and therefore, results should be considered with due caution, due to being somewhat underpowered. It is possible that the positive effects over time were due to either regression to the mean or natural recovery processes, rather than the impact of the chatbots. It is worth noting that, based on the power calculation, sufficient power was achieved for baseline to postintervention comparisons. Future studies comparing *chatbots* in clinical samples would, therefore, benefit from randomly allocating to a no treatment−passive control, to compare clinical outcomes for conversational agents against any natural recovery rate. Participants were recruited from an organization whereby membership would imply that they were open-minded to new experiences and willing to learn, and therefore, the results may not generalize to other older adults in terms of willingness to interact with a *chatbot*. It would also be useful to determine the average *chatbot* session length, when the time of the session is not recommended or limited or when there is a clinical problem being discussed.

The prompt concerning conversations needing to last approximately 20 min may have impeded deeper engagement, thus preventing problem resolution. In terms of future research, there are no published studies that investigate how the SUS interacts with other dimensions of e-therapy, such as treatment credibility, and further studies should examine this in more depth. Future studies should also assess clinical populations across the age ranges to evaluate if system usability and clinical outcomes differ between diagnoses. If the primary outcome is problem solving, then a conversational agent that follows the principles and stages of problem solving also needs to be developed and tested. The study would have benefited from a longer follow-up period, and future studies should enable short- and long-term follow-up. A possible innovation in future studies would be to adopt a patient preference trial methodology, whereby participants are offered the choice either MYLO or ELIZA (ie, to suit their preference) and those participants that are ambivalent about the choice of chatbot can be randomized.

Due to increasing referral pressure on mental health services, the flexibility of service delivery systems is important in reducing wait times for treatment, particularly in geographically remote regions. Approximately 5% to 15% of the older people also report chronic loneliness [42], and thus, *chatbots* appear to offer some potential in terms of offering conversational support to isolated older people. Talking with a conversational agent may also be particularly useful for psychological disorders involving high levels of shame and embarrassment. Indeed, the real utility of *chatbots* may be in supplementing traditional psychotherapies by reducing the number of sessions needed, because the conversational agent can provide between-session support and the therapist can focus on challenging change work during face-to-face treatment sessions. Similar models of augmenting face-to-face therapy with electronic alternatives have been discussed by Broglia et al [43]. The manner in which conversational agents could be usefully integrated into care pathways of routine psychological services needs to be explored.

Conclusions

In conclusion, this study sought to contribute to the evidence base regarding the utility and effectiveness of *chatbots* for psychological problems. This was achieved by comparing and testing 2 equivalent systems in terms of their acceptability, helpfulness, and effectiveness using a nonclinical older adult sample. The results have proven to be both similar and different from previous studies in working-age adults; MYLO is more helpful, but neither conversational agent differentially enabled problem resolution. Future controlled studies are clearly needed to further evaluate the clinical and health economic utility of conversational agents, but the context needs to be more clinical, outcomes need to be evaluated over longer periods, and system usability needs careful consideration.

Acknowledgments

This work was supported by a Doctor of Philosophy studentship awarded by the University of Sheffield to the first author MB and an Economic and Social Research Council grant (number ES/L001365/1).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Standard Protocol Items: Recommendations for Interventional Trials (SPRINT) diagram.
[PNG File , 139 KB - jmir_v22i5e16794_app1.png]

References

1. Richards D, Timulak L, Doherty G, Sharry J, Colla A, Joyce C, et al. Internet-delivered treatment: its potential as a low-intensity community intervention for adults with symptoms of depression: protocol for a randomized controlled trial. BMC Psychiatry 2014 May 21;14(1):147 [FREE Full text] [doi: 10.1186/1471-244X-14-147] [Medline: 24886179]
2. Kessler D, Lewis G, Kaur S, Wiles N, King M, Weich S, et al. Therapist-delivered internet psychotherapy for depression in primary care: a randomised controlled trial. The Lancet 2009 Aug 22;374(9690):628-634. [doi: 10.1016/S0140-6736(09)61257-5] [Medline: 19700005]

3. Bennion MR, Hardy GE, Moore RK, Kellett S, Millings A. e-Therapies in England for stress, anxiety or depression: how are apps developed? A survey of NHS e-therapy developers. BMJ Health Care Inform 2019 Jun;26(1):e100027 [FREE Full text] [doi: 10.1136/bmjhci-2019-100027] [Medline: 31171556]
4. Helgadóttir FD, Menzies RG, Onslow M, Packman A, O'Brian S. Online CBT I: Bridging the Gap Between Eliza and Modern Online CBT Treatment Packages. Behav Chang 2009 Dec;26(4):245-253. [doi: 10.1375/bech.26.4.245]
5. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
6. Crabb R, Cavanagh K, Proudfoot J, Learmonth D, Rafie S, Weingardt K. Is computerized cognitive-behavioural therapy a treatment option for depression in late-life? A systematic review. Br J Clin Psychol 2012 Nov;51(4):459-464. [doi: 10.1111/j.2044-8260.2012.02038.x] [Medline: 23078214]
7. Elsegood K, Powell D. Computerised cognitive-behaviour therapy (cCBT) and older people: A pilot study to determine factors that influence willingness to engage with cCBT. Couns Psychother Res 2008 Sep;8(3):189-192. [doi: 10.1080/14733140802163914]
8. Botella C, Etchemendy E, Castilla D, Baños RM, García-Palacios A, Quero S, et al. An e-health system for the elderly (Butler Project): a pilot study on acceptance and satisfaction. Cyberpsychol Behav 2009 Jun;12(3):255-262. [doi: 10.1089/cpb.2008.0325] [Medline: 19445633]
9. Zou JB, Dear BF, Titov N, Lorian CN, Johnston L, Spence J, et al. Brief internet-delivered cognitive behavioral therapy for anxiety in older adults: a feasibility trial. J Anxiety Disord 2012 Aug;26(6):650-655. [doi: 10.1016/j.janxdis.2012.04.002] [Medline: 22659078]
10. Dear BF, Zou J, Titov N, Lorian C, Johnston L, Spence J, et al. Internet-delivered cognitive behavioural therapy for depression: a feasibility open trial for older adults. Aust N Z J Psychiatry 2013 Mar;47(2):169-176. [doi: 10.1177/0004867412466154] [Medline: 23152358]
11. Spek V, Nyklícek I, Smits N, Cuijpers P, Riper H, Keyzer J, et al. Internet-based cognitive behavioural therapy for subthreshold depression in people over 50 years old: a randomized controlled clinical trial. Psychol Med 2007 Dec;37(12):1797-1806. [doi: 10.1017/S0033291707000542] [Medline: 17466110]
12. Titov N, Dear BF, Ali S, Zou JB, Lorian CN, Johnston L, et al. Clinical and cost-effectiveness of therapist-guided internet-delivered cognitive behavior therapy for older adults with symptoms of depression: a randomized controlled trial. Behav Ther 2015 Mar;46(2):193-205. [doi: 10.1016/j.beth.2014.09.008] [Medline: 25645168]
13. Spek V, Cuijpers P, Nyklícek I, Smits N, Riper H, Keyzer J, et al. One-year follow-up results of a randomized controlled clinical trial on internet-based cognitive behavioural therapy for subthreshold depression in people over 50 years. Psychol Med 2008 May;38(5):635-639. [doi: 10.1017/S0033291707002590] [Medline: 18205965]
14. Dear BF, Zou JB, Ali S, Lorian CN, Johnston L, Sheehan J, et al. Clinical and cost-effectiveness of therapist-guided internet-delivered cognitive behavior therapy for older adults with symptoms of anxiety: a randomized controlled trial. Behav Ther 2015 Mar;46(2):206-217. [doi: 10.1016/j.beth.2014.09.007] [Medline: 25645169]
15. International Organization for Standardization. 2018. ISO 9241-11:2018(en) Ergonomics of Human-System Interaction-Part 11: Usability: Definitions and Concepts URL: https://www.iso.org/obp/ui/fr/#iso:std:iso:9241:-11:ed-2:v1:en [accessed 2018-04-13]
16. Cavanagh K, Shapiro DA, van den Berg S, Swain S, Barkham M, Proudfoot J. The acceptability of computer-aided cognitive behavioural therapy: a pragmatic study. Cogn Behav Ther 2009;38(4):235-246. [doi: 10.1080/16506070802561256] [Medline: 19306147]
17. Vis C, Kleiboer A, Prior R, Bønes E, Cavallo M, Clark SA, et al. Implementing and up-scaling evidence-based eMental health in Europe: The study protocol for the MasterMind project. Internet Interv 2015 Nov;2(4):399-409. [doi: 10.1016/j.invent.2015.10.002]
18. Kleiboer A, Smit J, Bosmans J, Ruwaard J, Andersson G, Topooco N, et al. European COMPARative Effectiveness research on blended Depression treatment versus treatment-as-usual (E-COMPARED): study protocol for a randomized controlled, non-inferiority trial in eight European countries. Trials 2016 Aug 3;17(1):387 [FREE Full text] [doi: 10.1186/s13063-016-1511-1] [Medline: 27488181]
19. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
20. Fisk AD, Czaja SJ, Rogers WA, Charness N, Sharit J. Designing for Older Adults: Principles and Creative Human Factors Approaches. Boca Raton, Florida, United States: CRC Press; 2009.
21. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
22. Cavanagh K. Turn on, tune in and (don't) drop out: engagement, adherence, attrition, and alliance with internet-based interventions. In: Bennett-Levy J, Richards D, Farrand P, Christensen H, Griffiths K, Kavanagh D, et al, editors. Oxford Guide to Low Intensity CBT Interventions. London, UK: Oxford University Press; 2010:227.

23. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. Can J Psychiatry 2019 Jul;64(7):456-464 [FREE Full text] [doi: 10.1177/0706743719828977] [Medline: 30897957]
24. Rogers CR. A Way Of Being. Boston, Massachusetts, United States: Houghton Mifflin Harcourt; 1995.
25. Turkle S. Life On The Screen: Identity In The Age Of The Internet. New York, New York, United States: Simon & Schuster; 1997.
26. Carey TA. The Method Of Levels: How To Do Psychotherapy Without Getting In The Way. Hayward, California: Living Control Systems Publishing; 2006.
27. Powers WT. Behavior: The Control Of Perception. New Canaan, United States: Benchmark Publications; 1973.
28. Gaffney H, Mansell W, Edwards R, Wright J. Manage Your Life Online (MYLO): a pilot trial of a conversational computer-based intervention for problem solving in a student sample. Behav Cogn Psychother 2014 Nov;42(6):731-746. [doi: 10.1017/S135246581300060X] [Medline: 23899405]
29. Bird T, Mansell W, Wright J, Gaffney H, Tai S. Manage your life online: a web-based randomized controlled trial evaluating the effectiveness of a problem-solving intervention in a student sample. Behav Cogn Psychother 2018 Sep;46(5):570-582. [doi: 10.1017/S1352465817000820] [Medline: 29366432]
30. U3A: University of the Third Age. About URL: https://u3a.org.uk/about [accessed 2018-10-26]
31. Brooke J. SUS: A "Quick and Dirty" Usability Scale. Boca Raton, Florida, United States: CRC Press; 1996.
32. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum Comput Interact 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
33. Sauro J, Dumas JS. Comparison of Three One-Question, Post-Task Usability Questionnaires. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2009 Presented at: CHI'09; April 4-9, 2009; Boston, MA, USA p. 1599-1608. [doi: 10.1145/1518701.1518946]
34. Brooke J. SUS: a retrospective. J Usability Stud 2013 Feb;8(2):29-40 [FREE Full text]
35. Lovibond SH, Lovibond PF. Manual for the Depression Anxiety Stress Scales (Second edition). Sydney: Psychology Foundation of Australia; 1995.
36. CyberPsych. 2008. Eliza, The Computer Therapist URL: https://www.cyberpsych.org/eliza/ [accessed 2019-10-26]
37. Flückiger C, Del Re AC, Wampold BE, Horvath AO. The alliance in adult psychotherapy: a meta-analytic synthesis. Psychotherapy (Chic) 2018 Dec;55(4):316-340. [doi: 10.1037/pst0000172] [Medline: 29792475]
38. International Organization for Standardization. 2018. ISO 9241-11:2018(en), Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts URL: https://www.iso.org/obp/ui/fr/#iso:std:iso:9241:-11:ed-2:v1:en [accessed 2018-04-10]
39. Etzelmueller A, Radkovsky A, Hannig W, Berking M, Ebert DD. Patient's experience with blended video- and internet based cognitive behavioural therapy service in routine care. Internet Interv 2018 Jun;12:165-175 [FREE Full text] [doi: 10.1016/j.invent.2018.01.003] [Medline: 30135780]
40. de Wit J, Dozeman E, Ruwaard J, Alblas J, Riper H. Web-based support for daily functioning of people with mild intellectual disabilities or chronic psychiatric disorders: A feasibility study in routine practice. Internet Interv 2015 May;2(2):161-168. [doi: 10.1016/j.invent.2015.02.007]
41. van Ballegooijen W, Cuijpers P, van Straten A, Karyotaki E, Andersson G, Smit JH, et al. Adherence to internet-based and face-to-face cognitive behavioural therapy for depression: a meta-analysis. PLoS One 2014;9(7):e100674 [FREE Full text] [doi: 10.1371/journal.pone.0100674] [Medline: 25029507]
42. Pinquart M, Sorensen S. Influences on loneliness in older adults: a meta-analysis. Basic Appl Soc Psych 2001;23(4):245-266. [doi: 10.1207/s15324834basp2304_2]
43. Broglia E, Millings A, Barkham M. Counseling with guided use of a mobile well-being app for students experiencing anxiety or depression: clinical outcomes of a feasibility trial embedded in a student counseling service. JMIR Mhealth Uhealth 2019 Aug 15;7(8):e14318 [FREE Full text] [doi: 10.2196/14318] [Medline: 31418424]

Abbreviations

ANOVA: analysis of variance
CBT: cognitive behavioral therapy
DASS-21: depression, anxiety, and stress scales 21
e-therapies: electronic therapies
MOL: method of levels
MYLO: Manage Your Life Online
SUS: system usability scale
U3A: University of the Third Age

Edited by G Eysenbach; submitted 26.10.19; peer-reviewed by E Broglia, J Andrews, K Matsumoto; comments to author 16.11.19; revised version received 10.03.20; accepted 12.03.20; published 27.05.20.

Please cite as:
Bennion MR, Hardy GE, Moore RK, Kellett S, Millings A
Usability, Acceptability, and Effectiveness of Web-Based Conversational Agents to Facilitate Problem Solving in Older Adults: Controlled Study
J Med Internet Res 2020;22(5):e16794
URL: http://www.jmir.org/2020/5/e16794/
doi:10.2196/16794
PMID:32384055

Original Paper

The Effect of a Customized Nutrient-Profiling Approach on the Glycated Hemoglobin Levels of Patients With Type 2 Diabetes: Quasi-Experimental Study

Mayda Alrige[1], PhD; Riad Alharbey[2], PhD; Samir Chatterjee[3], PhD

[1]King Abdulaziz University, Jeddah, Saudi Arabia

[2]University of Jeddah, Jeddah, Saudi Arabia

[3]Claremont Graduate University, Claremont, CA, United States

Corresponding Author:
Mayda Alrige, PhD
King Abdulaziz University
6643 Abdullah Sulayman
Jeddah, 21589
Saudi Arabia
Phone: 966 12 695 2000
Email: malraegi@kau.edu.sa

Abstract

Background: Presently, dietary management approaches are mostly oriented toward using calorie-counting and diet-tracking tools that draw our attention away from the nutritional value of our food. To improve individuals' dietary behavior, primarily that of people with type 2 diabetes, a simple technique is needed to increase their understanding of the nutritional content of their food.

Objective: This study aimed to design, develop, and evaluate a customized nutrient-profiling tool called *EasyNutrition*. *EasyNutrition* was built to introduce the new concept of nutrient profiling by applying the Intelligent Nutrition Engine, an algorithm that we developed for ranking different food recipes based on their nutritional value. This study also aimed to investigate the efficacy of *EasyNutrition* in lowering glycated hemoglobin (HbA_{1c}) levels and improving dietary habits among people with type 2 diabetes.

Methods: We evaluated the utility of *EasyNutrition* using design science research in three sequential stages. This paper has elaborated on the third stage to investigate the efficacy of *EasyNutrition* in managing type 2 diabetes. A quasi-experimental study was conducted in a diabetes treatment center (n=28). The intervention group utilized *EasyNutrition* over 3 months, whereas participants in the control group utilized the standard of care provided by the center. Dietary habits and HbA_{1c} levels were measured to capture any change before and after experimenting with *EasyNutrition*.

Results: The intervention group (n=9) exhibited a statistically significant change between the pre- and postexposure results of their HbA_{1c} (t_9=2.427; P=.04). Their HbA_{1c} dropped from 8.13 to 6.72. This provided preliminary evidence of the efficacy of using a customized nutrient-profiling app in reducing HbA_{1c} for people with type 2 diabetes.

Conclusions: This study adds to the evidence base that a nutrient-profiling strategy may be a modern adjunct to diabetes dietary management. In conjunction with reliable dietary education provided by a registered dietician, *EasyNutrition* may have some beneficial effects to improve the dietary habits of people with type 2 diabetes.

(J Med Internet Res 2020;22(5):e15497) doi:10.2196/15497

KEYWORDS

mHealth; consumer health informatics; diet therapy; nutrient profiling; Hb A1c

Introduction

Problem Domain

Keeping up with a healthy, well-balanced diet is by no means easy. This is primarily challenging for patients with type 2 diabetes because it is related to maintaining a certain metabolic goal. The traditional dietary methods to manage diabetes include carbohydrate intake monitoring and *my plate*, among many others. Despite their effectiveness when followed strictly by patients, these methods are very cumbersome to deal with on a daily basis. The methods are based on specifying the quantity of certain nutrients. For example, following the *my plate* method, users must adhere to 25% protein, 25% grain and starchy foods, and 50% nonstarchy vegetables. For certain recipes, it is not practical to calculate those percentages. Dealing with quantities to manage diets is very tedious and time consuming.

Another problem with the current traditional dietary management methods is the focus on single nutrients instead of the overall nutritional value. Carbohydrate-restricted, fat-restricted diets are examples of these kinds of diets. The American Diabetes Association (ADA) released a new statement in 2013 summarizing the nutrition therapy recommendations for people with type 2 diabetes. As people eat food and no single nutrient such as carbohydrates, protein, and fat, this new statement designates a new section on eating patterns or plans. Patients with diabetes can still enjoy the food they like while, at the same time, keeping their diabetes under control. In addition, educating people about nutrition has to be aimed at modifying the factors that influence their dietary behavior rather than being aimed at increasing their knowledge about nutrition. Bader et al [1] found, through their comprehensive literature review, that knowledge-based nutrition education programs alone do not result in dietary behavior change. For instance, stating the recommended carbohydrate intake might increase one's knowledge about nutrition but not influence his or her dietary behavior to consume fewer carbohydrates. When this piece of information is applied in recipes and presented in a practical way, the chances are that patients would apply these tips and keep using them. This is the rationale behind behavioral nutrition, which is to change the target of nutrition education from the goal of increasing knowledge to that of modifying factors that influence dietary behaviors.

In addition, one more issue concerning the current dietary management tools is identified by the ADA in their new position statement on nutrition therapy recommendations. The ADA states that there is no single eating pattern that is best for everyone [2]. In their 2013 report, individualization is highlighted as a key element, as different people eat different food. When it comes to food choice, no solution is universal. Individual preferences contribute to one's eating plan. Not considering this aspect of individualization might result in individuals abandoning the dietary management tool.

Motivated by the aforementioned issues, we have developed a customized easy-to-follow dietary tool called *EasyNutrition*. *EasyNutrition* applies the concept of *nutrient profiling* in a mobile-based dietary app and aims to focus individuals' attention toward nutrition and raise their awareness about the nutritional value of the food recipes they choose. The engine behind *EasyNutrition* is a second artifact that we built to rank different food recipes, the *Intelligent Nutrition Engine*. This algorithm considers the 3 major macronutrients (ie, fat, protein, and carbohydrates), 2 micronutrients (ie, sodium and dietary fibers), and the number of calories the consumer needs daily. To present the nutritional value of a particular food recipe in a simple, easy-to-understand fashion, we adopted the notion of the traffic-light diet. The nutritional information is not given as a strict tricolor output. Rather, it is by analogy a color-coded food rating scale of eight values that rates the food recipe based on its nutrition from red (for extremely non-nutritious choices) to green (for optimal nutritious choices).

Background

Although genetics are an important consideration in health, during the past half-century, our genes have not measurably altered, and yet, we are significantly more overweight, obese, and prone to lifestyle-related diseases. As of 2014, more than one-third (36.5%) of US adults have obesity, according to the National Health and Nutrition Examination Survey data (2011-2014) [3]; 70.7% are overweight [4]; and 29.1 million Americans (9.3% of the population) are diabetics (as of 2012) [5]. The root of the problem of all these conditions is a poor diet. Tackling this problem is by no means easy and requires complex lifestyle changes. However, a healthy diet is a key component of a healthy lifestyle that can prevent the onset of chronic diseases or mitigate their severity [6].

Dietary Management

Managing diets is essential when it comes to diet-related chronic diseases. For healthy individuals, it is a preventative measure to maintain a healthy weight and facilitate overall well-being. Essentially, there are 4 different approaches to manage diets, summarized in a study by Arens-Volland et al [6]. These approaches encompass dietary recall, food records, self-management, and menu planning. In the first approach, the patient is asked over the phone about the amount of food and drinks typically consumed in a 24-hour period, along with the method of preparation and the brand of the food items. Using food records, patients will have the chance to do the same by writing down this information. In the self-management approach, a personal digital assistant software is used. The patient will set some predefined goals on what quantities of food to consume. In this case, patients are in charge of how to keep their diet under control no matter what kind of food is consumed. The last approach is menu planning where patients will have their meals planned based on previously identified food preferences. As the first 3 approaches are prone to the tediousness of counting and recording multiple times a day, the focus in this study was to utilize the last approach, which is menu planning.

Many studies have been conducted to develop and evaluate computerized dietary management approaches that are based on diet recall and food records [7,8]. These two methods are designed to both keep track of food intake and count calories or carbohydrates. These apps are promising to assist patients to better manage their diet. However, the notion of calorie counting is very tedious and entails many issues from the patients' perspective.

First, the underlying method, where users have to log their daily food intake, can suffer from the issue of recall. Examples of these food and calorie tracker apps include *MyFitnessPal*, *Lose It!*, and *Calorie Count*, among many others. These apps allow users to log their food daily, define personal weight loss goals, and review and analyze the gathered data against these goals. In this context, the second issue stems from a realistic limitation, which is the food database where users get to pick recipes from. The quality of the food tracking/dairy app is tightly dependent on this underlying database. However, even the largest current food databases are still far from being complete and often contain only country-specific products [9]. The notion of nutrition education or nutrition profiling helps alleviate these issues as it aims to educate patients about the underlying nutrients that constitute a good or bad food choice.

Nutrition Education

To better manage diets and sustain a healthy lifestyle, one has to be aware and knowledgeable about the nutritional content of the food consumed. Being educated and aware of macronutrition and micronutrition would contribute to one's overall health. This knowledge leads to a healthy behavior, which, in turn, holds the promise of preventing the onset of chronic diseases or mitigating their severity. Individualized diet education is effective to both understand diet requirements and control body weight and blood sugar levels [10]. This is especially true when such educational material is delivered in an easy-to-follow and a convenient-to-understand manner. However, knowledge-based nutrition education alone does not change dietary behavior. Behavioral nutrition education is more effective in changing dietary behavior. The effectiveness of the behavioral nutrition approaches was investigated by Bader et al [1]. They conducted a pilot study where they investigated one of the most preferred dietary management approaches: menu planning [11]. The study targeted patients with type 2 diabetes. The authors conducted a single-arm clinical trial to evaluate one of the commercial internet-based menu planning tools. They examined pre- to postintervention changes in body weight, blood pressure, and glycemia among overweight patients with type 2 diabetes mellitus (T2DM; n=33). Nutritional recommendations were operationalized into weekly internet-delivered menu plans. The findings indicated that there was 5% weight reduction in 0.18% of the participants. The study highlighted the effectiveness of the behavioral nutrition approaches that are less structured. These approaches include gaining familiarity with general nutrition principles, acquiring general planning frameworks (eg, carbohydrate exchanges, *points*), and, finally, planning tools to use meal replacements or help prepare meals [1].

Nutrient Profiling

Nutritional profiling aims to rank food based on their nutritional quality. It is driven by the focus on food quality instead of quantity. Individuals who follow high-score food choices would most likely improve their dietary behavior.

Simplicity is key when it comes to presenting nutritional information. It has been suggested that nutritional information on mobile phones should be easy to read and understand. Arsand et al [12] stated that approximate information is better than accurate facts that are difficult to understand but more precise. It was further suggested that the food information on phones for T2DM should not be too fine-grained because too much detailed information may result in user discouragement and little user friendliness.

This approach emphasizes the idea that carbohydrate and calorie intake counting has become less preferred in favor of more generalized nutritional information about the quality of food compositions. As noted, nutrient profiling is defined as the science of ranking foods according to their nutritional composition for reasons related to preventing disease and promoting health, as stated by the World Health Organization. This idea has led to the creation of many nutritional rating systems.

Driven by the idea of nutrition profiling, the traffic-light diet was developed by Epstein et al in the 1970s [13]. In this dietary approach, Epstein et al [13] used a tricolor palette to create an easy-to-follow diet for overweight children. The notion of a traffic-light diet had inspired new research for two decades because of its groundbreaking nature. The traffic-light diet is a structured eating plan that divides food by the color of the traffic signals. Green (go) is for low-calorie food that can be eaten at any time, orange (caution) is for moderate-calorie food that can be eaten occasionally, and red (stop) is for high-calorie food that should be eaten rarely. Since it was launched, pediatricians have widely used the traffic-light diet to encourage healthy eating habits (EH) among their patients.

Many studies have been conducted utilizing the *traffic-light* dietary approach and have shown promising results. The traffic-light diet is used as a part of a comprehensive treatment, and the results show a significant decrease in obesity in preadolescent children [14,15]. Significant changes in eating patterns have been reported when comprehensive obesity treatment has been combined with the traffic-light diet [7,13]. Reductions in *red foods* have been observed after treatment, with significant associations between changes in intake of *red foods* and weight loss [7] or a decrease in the overweight percentage [13].

In this study, we adopted the presentation approach of the traffic-light diet to present the nutrients contained in a food recipe. However, it was not a strict tricolor output. Rather, it was a color-coded food rating scale of eight values as it takes into consideration five different nutrients and not only the caloric count. It scales food based on its nutritional quality from red, for extremely unhealthy choices, to green, for optimal healthy choices, through intermediate colors.

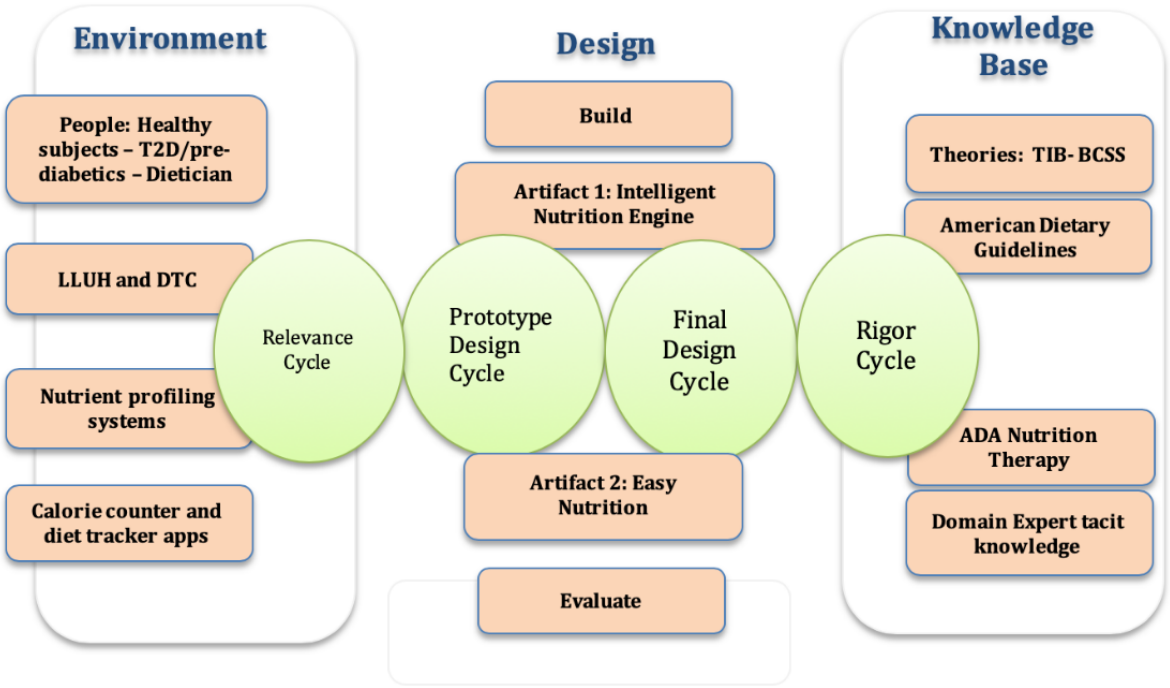
This study aimed to provide a nutrition educational tool to help people with type 2 diabetes learn about the nutritional content of the food they eat and, hence, improve their dietary behavior by choosing healthier, more nutritional recipes.

Research Approach

This study followed the design science research (DSR) approach suggested by Hevner and Chatterjee [8]. Hevner et al [16] presented design science as a legitimate research paradigm to be employed in the Information Systems research projects, where the goal was to solve practical problems. DSR aims to build and evaluate technology artifacts in an iterative process.

This process can be seen as the embodiment of 3 related cycles of activities: the relevance cycle, the rigor cycle, and the design cycle. The relevance cycle initiates the DSR project by addressing the contextual environment of the research project. The rigor cycle defines the knowledge base of the scientific foundations in the form of theoretical models/theories, experience, and expertise. It aims to ground the DSR artifact to the knowledge base and define its novelty. The design cycle iterates between the essential activities of build and evaluate. The 3 main cycles of DSR are illustrated in Figure 1.

Figure 1. Design science research methodology.



First Design Artifact: EasyNutrition

EasyNutrition is a customized dietary tool. This means that each user will get cuisine-driven recipes based on his/her own food preference. Once users download the *EasyNutrition* app, they are prompted to sign in or sign up (Figure 2). For new users, *EasyNutrition* will calculate their basal metabolic rate (BMR) based on age, gender, height, and weight. BMR is an estimate of how many calories the user would burn if he/she were to do nothing but rest for 24 hours. In other words, BMR is an estimation of the daily needed calories. In the profile, the user can edit favorite cuisines. These preferences will be saved in the user's profile and can be easily updated every time the user uses *EasyNutrition*.

Using *EasyNutrition's* Home page, users can find web-based recipes (tailored to their food preferences as well as their individual nutritional needs), browse the nutritional value of a variety of food recipes, or plan weekly meals. This meal plan can be based on either their favorite recipes (chosen earlier) or their favorite cuisine (Figure 3). In both cases, the resulting recipe will be displayed along with its overall nutritional value. The nutritional value is presented in a traffic-light scale, as can be seen in Figure 4, ranging from red (for a relatively poor nutrition choice) to green (for a relatively excellent nutrition choice) through some intermediate colors.

This scale will give the user an initial indication of how nutritious the chosen recipe is. If the user is interested to know more about the nutrients that lower the overall nutritional quality, he/she can click on the *nutrition* tab to find out which nutrient is beyond the recommended range, as can be seen in Figure 5. Apart from the nutritional quality, the pertinent information for each recipe is presented. This includes the ingredients, instructions, and some healthy tips on how to maximize the nutritional value of the selected recipe.

As the user saves and adds food recipes to his/her meal plan, he/she will be given either a gold, silver, or bronze trophy based on how active they are with the app. The app will categorize the user into one of the three different groups based on the number of recipes he/she adds to his/her meal plan.

Figure 2. EasyNutrition Sign up screen.

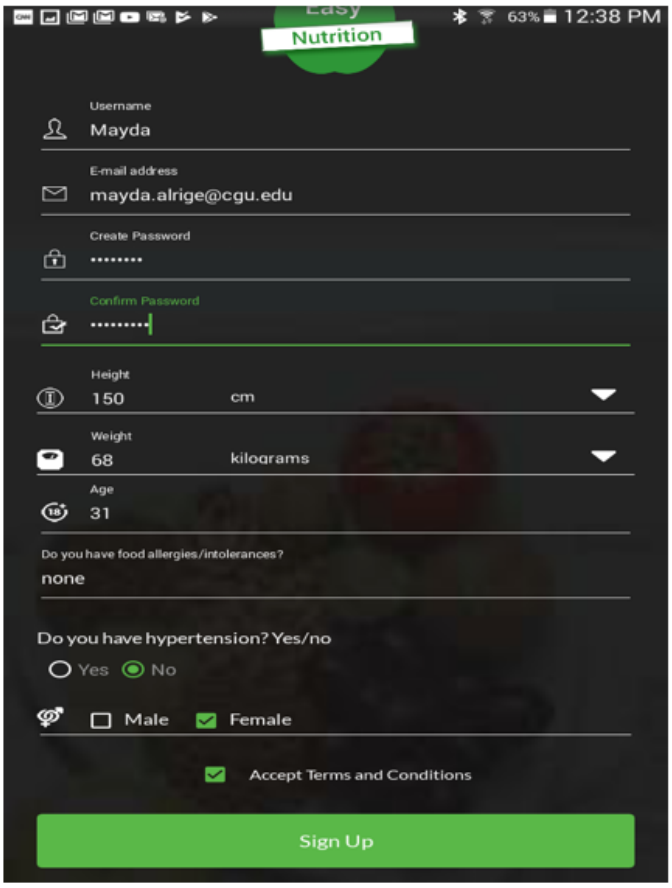


Figure 3. EasyNutrition Home page.

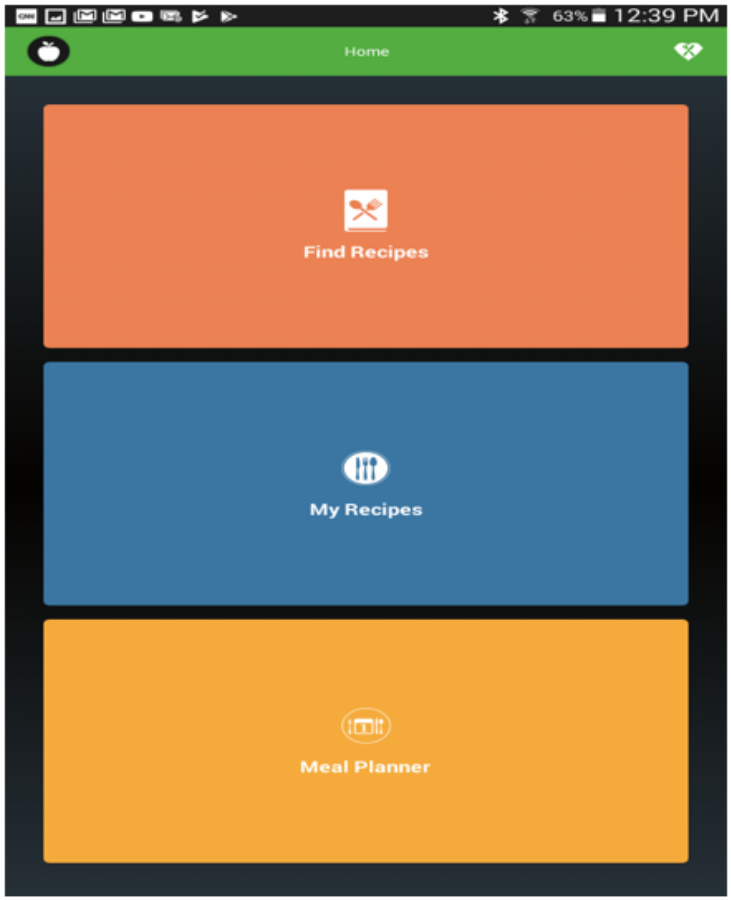


Figure 4. The overall nutritional value of the “Falafel salad” recipe.

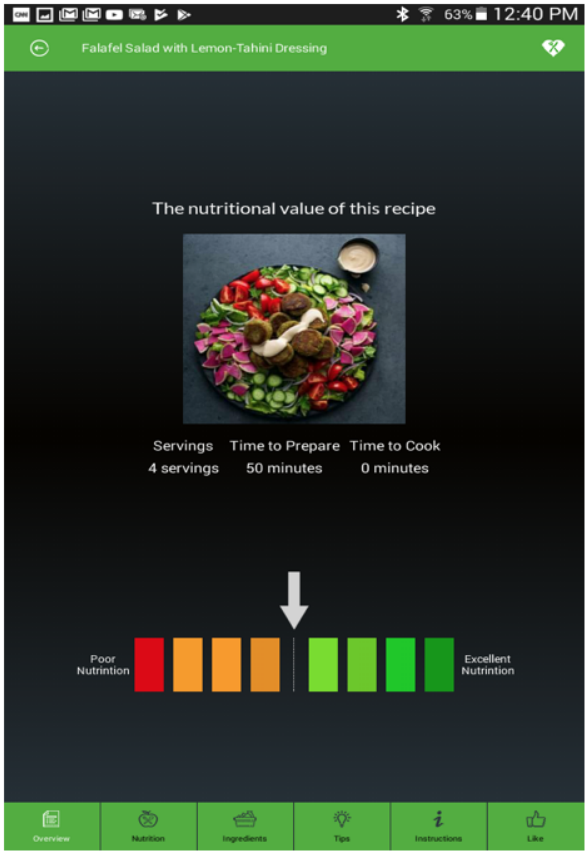
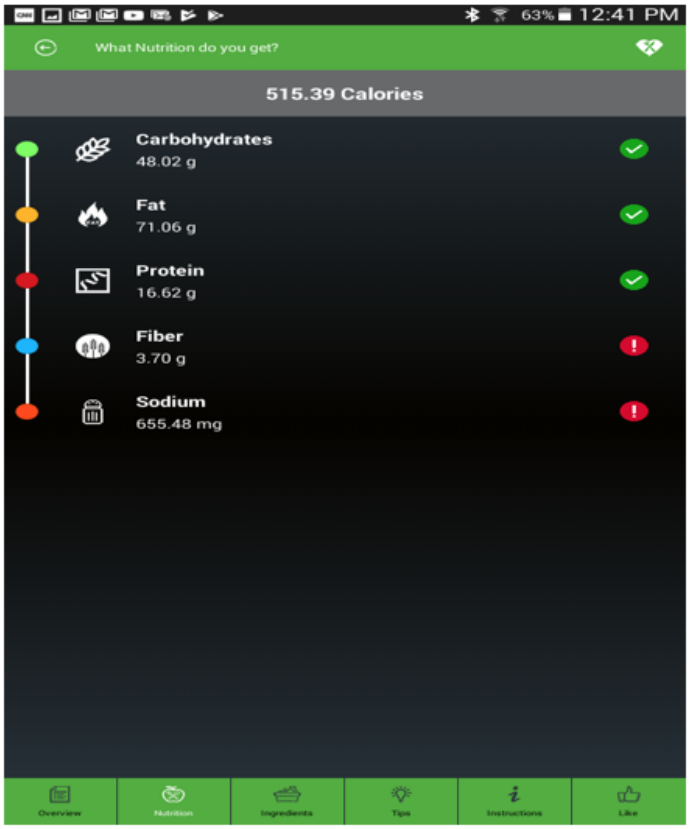


Figure 5. Nutrition break down of the recipe.



Second Design Artifact: The Intelligent Nutrition Engine

We developed the *Intelligent Nutrition Engine*, which is a food-ranking algorithm that encompasses 5 different nutrients: the 3 major macronutrients that derive calories (ie, fat, protein, and carbohydrates) and 2 micronutrients (ie, dietary fibers and sodium). The algorithm checks the calories, the percentage of macronutrients, and also the number of micronutrients in a selected recipe. These nutrients determine a healthy or an unhealthy choice. For people with diabetes, we consulted the nutrition therapy recommendations by the ADA [2] to find the important nutrients that have to be considered, along with the appropriate limit for each nutrient. Some of these criteria have slightly different specifications for comorbid type 2 diabetes and hypertension. For example, the sodium recommendation for the general population is less than 2300 mg/day. However, for individuals with both diabetes and hypertension, a further reduction in sodium intake should be individualized (no more than 1500 mg/day). In addition, for comorbid T2DM and hypertension, no more than 7% of fat should come from saturated fat. These criteria were reviewed and validated by the registered dietician from Loma Linda University Medical Center. A list of these nutrients is outlined in Table 1.

Harris-Benedict Equation

All these nutrients' amounts/percentages are based on one's daily calorie intake. As the recommended calorie intake differs based on age, gender, height, and weight, we applied the Harris-Benedict equation to determine the recommended calorie intake or the BMR based on these factors [17]. The final formula encompassing all these nutrients as well as the calories will produce a single holistic number that represents the overall nutritional quality of a particular recipe. The detailed steps of the algorithm can be seen in Textbox 1. For the sake of simplicity, we presented this number in a traffic-light scale that ranges from red to green through some intermediate colors.

The cursor would start right in the middle of the traffic-light scale as an initial score for any given food recipe, as can be seen in Figure 6. The cursor would move to the right if a certain nutrient is within the recommended percentage/amount. The cursor would move to the left if a certain nutrient exceeds the maximum limit or fails to meet the minimum limit of the recommended range. This algorithm was applied to Spoonacular, the largest web-based food Application Programming Interface (API), to rank different food recipes accordingly.

Table 1. The nutrition limits per daily recommended percentages/amounts of total calorie intake.

Nutrients	Intake
Macronutrients	
Carbohydrates	• 45%-65%
Fats	• 25%-35% • For those with hypertension, no more than 7% of this percentage should come from saturated fat
Protein	• 15%-20%
Micronutrients	
Dietary fibers	• 20-30 grams
Sodium	• No more than 2300 mg • No more than 1500 mg daily for diabetics who have hypertension

Textbox 1. The algorithm logic to rank food recipes.

- Calculate the basal metabolic rate (BMR) based on gender, age, height, and weight based on the Harris-Benedict equation
- Divide the number by 3 assuming 3 meals a day, to get the ideal recipe calories for a particular user
- Recommended recipe calories (RRC)=BMR/3
- Obtain the chosen recipe calories (CRC) and compare the value with RRC

If CRC≤RRC, and is within the recommended range, CRC_w=1, otherwise CRC_w=−1

- Obtain the amount of carbs in grams
- Convert this amount into calories (1 gram of carbs provides 4 calories)

0.45% *Rc*≤*C*≤0.65% *Rc*, where *C* is carbs

If it is within the recommended range, C_w=1, otherwise C_w=−1; if *C*<0.45%, C_w=0

- Obtain the amount of fats in grams
- Convert this amount into calories (1 gram of fat provides 9 calories)

0.25% *Rc*≤*F*≤0.35% *Rc*, where *F* is fat

If it is within the recommended range, F_w=1, otherwise F_w=−1; if *F*<0.25%, F_w=0, show the message: *Good job. This choice is low in fat*.

If the patient with diabetes has hypertension, then another test for saturated fat will be conducted: fat sugars should not exceed 0.07% of the total fat

- Obtain the amount of protein in grams
- Convert this amount into calories (1 gram of protein provides 4 calories)

0.15% *Rc*≤*Pr*≤0.20% *Rc*, where *Pr* is protein

If it is within the recommended range, Pr_w=2, otherwise Pr_w=−2; if *P*≤15% R_c, P_w=0

- Obtain the amount of sodium in milligrams

S≤450, where *S* is sodium

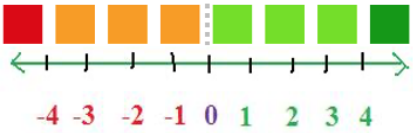
If *S*≤450, S_w=1; if *S*≥450, S_w=−1

- If the patient with diabetes has hypertension, then the amount has a lower cut-off value (<1500/3 mg)
- Obtain the amount of dietary fibers in grams

6 *g*≤*Df*≤10 *g*, where *Df* is dietary fibers

If *Df* is within the recommended range, Df_w=2; if *Df*>10, Df_w=3; if *Df*<6, Df_w=0

Figure 6. The nutritional score presented behind a traffic-light scale.

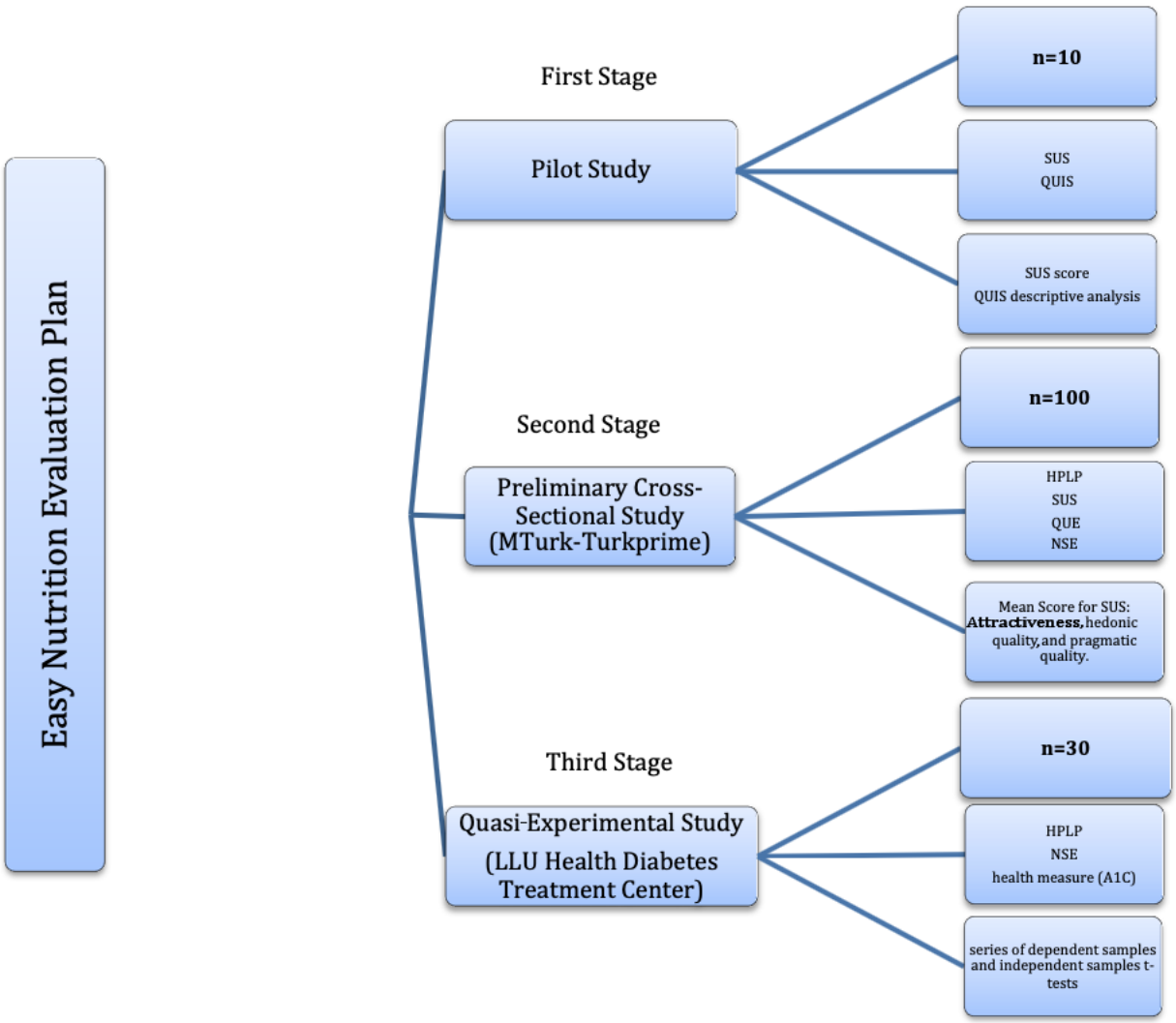


Methods

Evaluation Plan

During this study, we evaluated the utility, efficacy, and quality of the app, *EasyNutrition*, from a sociotechnical perspective in three sequential stages (Figure 7). The first stage was a pilot study to test the usability and understandability of *EasyNutrition*'s interfaces. The results can be found in a published paper by Alrige et al [11]. Once we established that the technology is sufficient to use (usable and understandable), we tested the full version of *EasyNutrition* on a wider population in the second stage. The goal was to evaluate *EasyNutrition* for its functionality and quality as a customized dietary tool, and the results were published in a paper by Alrige and Chatterjee [18]. Once we established the usability, quality, and content validity of *EasyNutrition* as a customized dietary tool to help users improve their food choices, we implemented the third stage to test *EasyNutrition*'s efficacy in managing type 2 diabetes. Details about the subjects, study type, measurements, and analysis can be seen in this paper that elaborates on the third stage of the evaluation where we conducted a quasi-experimental field study. In particular, we employed the two-group pretest-posttest design recommended for medical informatics studies to introduce a new intervention [19]. A pre-post intervention study allowed us to capture any change in participants' dietary behavior/attitude before and after experimenting with *Easy Nutrition.* Both Claremont Graduate University (Institutional Review Board [IRB] number 2964) and Loma Linda University (IRB number 5170397) provided an IRB approval, as we conducted subject recruitment in one of Loma Linda University's health facilities.

Figure 7. The evaluation plan (subjects, measures, and analysis).



Setting and Sampling

Subjects were recruited from the diabetes treatment center (DTC), Loma Linda University Health. The center offers classes on diabetes education in 4 different sessions: the first and second sessions are on nutrition, the third one is on physical activity, and the last one is after 2 to 3 months of follow-up with the patients to provide a plan for them to move forward and live with diabetes. Diabetes educators and registered dieticians facilitate those sessions. We attended the sessions to introduce *EasyNutrition* and recruit patients after obtaining their consent. We attended seven classes in the DTC for recruitment, starting from December 20, 2017, and ending on January 10, 2018. The classes were offered during the morning (9 AM-11 AM), afternoon (1:30 PM-3:30 PM), and evening (5:30 PM-7:30 PM) sessions. We followed up with these cohorts of patients during the study course of 3 months before their fourth and last class at the end of their DTC session. The patients were in five different cohorts. The total number of participants was 28, with 5 patients with prediabetes and 23 patients with type 2 diabetes. In total, 12 patients were signed up for the intervention group (group number 1), and 16 patients were signed up for the control group (group number 2). The sample size (n=28) was determined for a two-group, pretest, posttest intervention based on the following test parameters: statistical power=0.80, (medium) effect size=0.5, and significance level=.05.

Procedures and Measures

In the first class, we started the recruitment by introducing ourselves and our study and explained how relevant it is to their diabetes education session. This included playing a 2-min short demo about *EasyNutrition*. Upon their agreement to participate, subjects completed the following activities.

First, as part of the DTC standard of care (SOC), subjects' blood was drawn to measure their blood sugar (glycated hemoglobin, HbA_{1c}), and this was self-reported to the research personnel. The HbA_{1c} is measured in units of mmol/mol. The HbA_{1c} test measures how much hemoglobin in the blood has become glycated (chemically bonded with glucose).

Second, researchers assigned participants either to the intervention group (group number 1) or to the control group (group number 2), based on the kind of smartphone he/she had (patients with an Android-based smartphone were assigned to group number 1).

Third, participants answered a quick questionnaire about their EH using the Health-Promoting Lifestyle Profile (HPLP) nutrition subscale. The administered questionnaire was adapted from the HPLP, a 48-item questionnaire of self-reported health-promoting lifestyle habits. HPLP consists of 6 scales: self-actualization, health-responsibility, exercise, nutrition, interpersonal support, and stress management. The final structure of the 48-item HPLP was found to have high internal consistency with an alpha coefficient of .922. Sets of items assigned to each of the 6 factors were examined for their reliability as subscales. The nutrition subscale is found reliable with an alpha coefficient equal to .757.

We utilized the Nutrition subscale, a 10-item, 4-point Likert scale that ranges from Never (1), Sometimes (2), Often (3), to Routinely (4) to indicate how often an individual engages in each behavior. The minimum score was 10, and the maximum score was 40 [20].

On the basis of participants' answers to these questions, we generated a score reflecting their EH. This score was used as a baseline for postcomparison for both groups. An increase in this score suggests improvement in the patient's dietary behavior. This can establish a potential association between the intervention (interacting with *EasyNutrition*) and health outcomes.

Intervention

If the patient was assigned to the intervention group, he/she was asked to download the app, *EasyNutrition*, on his/her Android-based phone or tablet. We trained the patients to use the app. Patients were shown how to sign up by entering their basic demographic information (age, height, weight, and gender). They were introduced to the concept of BMR, which is the daily calories their bodies need. They were told that all the nutritional value of the recipes would be based on this BMR value. This is the first form of customization. The second and more important form of customization has been presented by suggesting recipes tailored to the individuals' favorite cuisines. They were shown how to find web-based recipes and learn about their nutrition breakdown and the quality of ingredients. They were taught how to add the recipes to their meal plan and how to track their usage in their profiles. Every 2 weeks, they received either a text message or a phone call to ask questions about their app usage. Over 3 months, patients in the intervention group used *EasyNutrition* every other week to learn about the nutrition breakdown as well as the nutritional value of a variety of food recipes tailored to their food preferences. Patients in the control group continued receiving the SOC treatment offered by the DTC.

Fourth, at the end of the study, all subjects filled in the same questionnaire about their EH. In addition, their HbA_{1c} was measured again to compare it with the benchmark data from the presurvey data. The follow-up classes at the DTC we attended were at the following dates/times: Monday, February 26, 2018 (5:30-7:30 PM); Wednesday, February 28, 2018 (9 AM-11 AM); Monday, March 26, 2018 (1:30 PM-3:30 PM); and Wednesday, March 28, 2018 (5:30 PM-7:30 PM). One class was designated for only those with prediabetes, and it was offered under the diabetes prevention program. This was on Wednesday, March 14, 2018 (5:30 PM-7:30 PM).

In addition, all subjects were asked to fill in the nutrition self-efficacy (NSE) questionnaire [21]. On the basis of the health belief model [21], the self-efficacy construct has an effect on engaging in health-promoting behavior [22]. Self-efficacy is defined as one's ability or confidence in making a behavior change. The measurement of patients' NSE was considered to account for habitual unhealthy behaviors, such as overeating, and test any confounding factors that might have affected their postmeasurement data.

Analysis

We compared the mean of both eating habit composite scores and HbA_{1c} before and after the study for both groups

(intervention and control). Thus, a series of independent and dependent sample one-tailed *t* tests were conducted to test if there was any statistically significant difference within and between groups in terms of patients' HbA_{1c} and dietary habits. A positive change indicates causality between the intervention and dietary behavior. This would provide preliminary evidence that a potential behavior change might have occurred because of the intervention.

Results

Participants Characteristics

Subjects are patients with either type 2 diabetes or prediabetes. Out of the 28 patients who signed up in the third stage, 21 (75%) completed their participation. Of these, 9 patients were men and 12 were women, with the majority being white (11 out of 21). To measure *EasyNutrition* efficacy, we evaluated the subjects' EH as well as their HbA_{1c} before and after the experiment for both groups.

Pre-Exposure Analysis

Participants in both groups (9 subjects in the intervention group and 12 subjects in the control group) were administered the presurvey to determine the pre-exposure composite scores of their EH and, hence, establish a benchmark healthy score. The mean composite score for EH was 23.33 for the intervention group and 25.83 for the control group. As for the HbA_{1c}, the mean was 8.13 for the intervention group and 7.13 for the control group (Table 2). To ensure that the means of the 2 groups were not significantly different at the start, an independent samples *t* test was conducted on both the composite EH score (t_{21}=1.27; *P*=.22) and HbA_{1c} (t_{21}=−0.967; *P*=.31). The results in Table 3 show that there is no statistically significant difference between the 2 groups. This illustrates that, before the beginning of the intervention, the difference of the mean between the 2 groups was indeed not significantly different.

Table 2. Descriptive statistics on the eating habits composite score and glycated hemoglobin before and after the experiment.

Group	Values	
	Mean (SD)	SE mean
Intervention (n=9)		
HbA_{1c}[a]		
Pre-HbA_{1c}	8.13 (2.83)	0.94
Post-HbA_{1c}	6.72 (1.27)	0.42
EH[b]		
Pre-EH	23.33 (3.08)	1.02
Post-EH	24.33 (3.54)	1.18
Control (n=12)		
HbA_{1c}		
Pre-HbA_{1c}	7.13 (1.49)	0.43
Post-HbA_{1c}	6.90 (1.64)	0.47
Pair 2		
Pre-EH	25.83 (5.27)	1.52
Post-EH	26.08 (2.94)	0.85

[a]HbA_{1c}: glycated hemoglobin.

[b]EH: eating habits.

Table 3. Pre-exposure independent *t* test.

Group	Levene test		*t* test for equality of means			
	F (df=20)	Significance (*P* value)	*t* test (df=19)	Significance (two-tailed; *P* value)	Mean difference	SE difference
Eating habits	3.39	.08	1.27	.22	2.50	1.98
Glycated hemoglobin	6.39	.02	−1.05	.31	−1.00	0.95

Postexposure Analysis

Both groups exhibited an increase in their composite dietary habits score in the postexposure results and a decrease in HbA_{1c} (Table 4). To objectively measure the changes in the means of the 2 groups, an assessment of statistical significance was conducted through a series of dependent and independent samples *t* tests.

First, a paired samples *t* test was conducted to assess if the changes within the groups are considered statistically significant. The first analysis consisted of two dependent *t* tests that were conducted for each group, individually. The dependent variables are HbA_{1c} and the composite score of EH. As for HbA_{1c} in the control group (n=12), the HbA_{1c} mean dropped from 7.13 to 6.90. This change in the HbA_{1c} was very minimal and not statistically significant (t_{12}=1.30; *P*=.22). The intervention group (n=9), however, exhibited a statistically significant change between the pre- and postexposure results of their HbA_{1c} (t_9=2.42; *P*=.04). Their HbA_{1c} had dropped from 8.13 to 6.72 (Table 5).

In addition, the second analysis consists of two independent samples *t* tests to assess the change between the 2 groups after the experiment in terms of both their EH composite score (EH) and HbA_{1c}. The results of the independent samples *t* tests showed a statistically significant difference in the postresults between the control and intervention groups in terms of their HbA_{1c} (t_{21}=1.94; *P*=.04). However, the difference between the 2 groups in terms of their EH was not statistically significant (t_{21}=−.456; *P*=.65; Table 6).

Table 4. Changes in both eating habits composite score and glycated hemoglobin.

Group	Before	After
Intervention		
HbA_{1c}[a]	8.10	6.60
EH[b]	23.3	24.4
Control		
HbA_{1c}	7.10	6.90
EH	25.9	26.10

[a]HbA_{1c}: glycated hemoglobin.

[b]EH: eating habits.

Table 5. Dependent samples *t* test on both glycated hemoglobin and eating habits for both groups.

Group	Values			
	Mean (SD)	SE mean	*t* test (df)	Significance (*P* value)
Intervention group (n=9)				
Pre-HbA_{1c}[a] to post-HbA_{1c}	1.41 (1.75)	0.58	2.42 (8)	.04
Pre-EH[b] to post-EH	−1.00 (2.74)	0.91	−1.09 (8)	.31
Control group (n=12)				
Pre-HbA_{1c} to post-HbA_{1c}	0.23167 (0.62)	0.18	1.30 (11)	.22
Pre-EH to post-EH	−0.25000 (4.31)	1.24	−0.20 (11)	.84

[a]HbA_{1c}: glycated hemoglobin.

[b]EH: eating habits.

Table 6. Between-group differences for both glycated hemoglobin and eating habits.

Group	Levene test		*t* test for equality of means			
	F (df=20)	Significance (*P* value)	*t* test (df=19)	Significance (two-tailed; *P* value)	Mean difference	SE difference
Glycated hemoglobin difference	12.67	.002	1.94 (19)	.04	1.18	0.54
Eating habits difference	3.30	.09	−0.456 (19)	.65	−0.75	1.64

Discussion

Practical Implications

The quasi-experiment conducted for this study showed that there is a statistically significant difference within and between subjects in terms of their HbA_{1c}. This provides strong preliminary evidence about the efficacy of using a nutrient profiling–based dietary app to present the nutritional value of different food recipes in a customized meal plan. Combined with the nutrition classes offered by DTC, interacting with *EasyNutrition* was effective in lowering HbA_{1c} in patients with diabetes by focusing their attention on nutrition. This is in line with the findings of Bader et al [1] that customized menu planning is one of the most effective methods for dietary management compared with other methods. The customized menu plans provided by *EasyNutrition* in this study enable patients with type 2 diabetes to learn about nutrition and nutritional value based on what they normally eat.

One might argue that the nutrition classes offered by DTC do have an effect on patients in lowering their HbA_{1c}. However, we tried to counterbalance this potential effect by having a control group. Obtaining pretest measurements on both the intervention and control groups allows us to assess the initial comparability of the groups. The assumption is that if the intervention and the control groups are similar at pretest, there is a smaller likelihood of important confounding variables differing between the 2 groups [19]. In addition, we measured the NSE for both groups to account for any habitual unhealthy behavior. The mean of the NSE was 15.25 for the control group and 14.67 for the intervention group (Table 7). The difference was negligible (t_{21}=0.49; *P*=.63 as can be seen in Table 8) and indicated that there is a little chance that NSE would have been a confounding factor in this study.

Table 7. Nutrition self-efficacy descriptive statistics.

Type	Values	
	Mean (SD)	SE mean
Nutrition self-efficacy		
Control (n=12)	15.25 (3.22)	0.93
Intervention (n=9)	14.67 (1.66)	0.55

Table 8. Nutritional self-efficacy between-group difference.

Type	Levene test		*t* test for equality of means				95% CI of the difference
	F (df=20)	Significance (*P* value)	*t* test (df=19)	Significance (two-tailed; *P* value)	Mean difference	SE difference	
Nutritional self-efficacy	2.06	.17	0.49	.63	0.58	1.18	−1.89 to 3.06

For patients with diabetes attending the DTC or any other diet management center, keeping up with a healthy nutrient-dense diet has always been a challenge. According to personnel from the DTC, most patients fall back into their old dietary habits as soon as they leave the center. By having a customized dietary app, such as *EasyNutrition*, in hand, the chances are greater that they will stay motivated enough to adhere to a healthy regimen. This is primarily true especially if the dietary app is presenting the nutrition information in a simple manner that does not interfere with day-to-day activities. Nutrient profiling as utilized in *EasyNutrition* has shown to be beneficial in this regard.

Research Contribution

This study contributes to the body of knowledge on 2 broad levels: to society and to science. The first level of contribution is a new method for nutrition education. The *Intelligent Nutrition Engine* was developed considering 5 major nutrients (3 major macronutrients and 2 micronutrients) to underpin the nutrient-profiling food rating scale. The *Intelligent Nutrition Engine* classifies different food recipes based on their nutritional value in the form of a traffic-light scale. In fact, nutrition education accounts for a large proportion of diabetes management or the management of any other diet-related chronic disease [6]. Thus, it is important to find a simple, intuitive, easy-to-understand mechanism to present nutritional values. The algorithm developed in this study is based on the notion of behavioral nutrition approaches stressed by Bader et al [1]. The notion behind behavioral nutrition approaches is to increase the likelihood that participants will implement the strategy learned. *EasyNutrition* does not offer general nutritional information. Rather, the nutritional value of the food recipe is presented in a simple, easy-to-understand manner. This study adds to the evidence base that a nutritional behavior strategy may be a modern adjunct to diabetes dietary management. In conjunction with reliable dietary education provided by a registered dietician, the app may have some beneficial effects to improve the dietary behavior of patients with diabetes.

In addition, the steps of the algorithm can be viewed as a set of design principles. The algorithm can be tailored to tackle different health conditions. Both the nutrients and the criteria for each nutrient can be tailored according to the health condition being treated. For example, cardiovascular diseases (CVDs) have certain nutrition therapy recommendations that are slightly different than those for patients with diabetes.

Diabetes gives priority to carbohydrate consumption, whereas CVDs give special attention to fat consumption.

The artifact, *EasyNutrition*, signals the second level of contribution. Using DSR, *EasyNutrition* was rigorously developed as a customized, nutrient-profiling tool. *EasyNutrition* applies the Intelligent Nutrition Engine to present the nutritional value of different food recipes in a simple manner. This nutritional value is better absorbed if it is tailored to individual preferences. According to Triandis [17] in his interpersonal behavior theory, habits do influence behavior. Thus, including individual habitual preferences would increase the chance to change dietary behavior toward a healthier one. The design process of *EasyNutrition* was delineated thoroughly through a new DSR framework. This includes design requirement extraction and design feature determination in light of the main design principles followed: customization and simplicity.

Limitations and Future Work

Although the results of this study are promising, there are some limitations that allow room for further improvement and map out new directions for future research. The first set of limitations concerns the design artifact, *EasyNutrition*, whereas the second set of limitations concerns the Intelligent Nutrition Engine. First, regarding the design of *EasyNutrition*, the app presents the overall nutritional value in a traffic-light scale to promote simplicity. For people with color blindness, this presents an issue as they cannot see the differences of the gradients of the three main colors of the traffic light. Another presentation mechanism can be added above or below the traffic light to mitigate this issue. A simple presentation mechanism such as numerical values or emoticon representations can be utilized. Second, the nutritional value of every recipe does not allow ingredient substitution, that is, if the user ended up cooking one recipe but chose to substitute, for example, white rice with brown rice or decided to grill the chicken breast instead of frying it, the nutritional value is still the same. Dynamic nutritional value calculation would allow users to substitute ingredients and then adjust the overall nutritional value accordingly.

The second set of limitations concerns the Intelligent Nutrition Engine artifact. This algorithm is inspired by the novel concept of *NuVal*, a system that has its own dataset and covers almost all the food products/packaged meals. Owing to financial limitations, we were not able to consult the NuVal dataset or API and utilize the NuVal scores. A good research direction would be to utilize NuVal scores that consider about 30 nutrients and apply them to rank different food recipes. This would give a more comprehensive dietary tool that can possibly present the nutritional value not only for food recipes but also for food products, restaurant meals, and many others.

The third set of limitation concerns the study design. In the third stage of this study, we conducted a quasi-experiment to evaluate the effect of using *EasyNutrition* on managing type 2 diabetes. The sample size was borderline small for the statistics used (n=28). To generalize the results for the effect of using *EasyNutrition* on reducing HbA_{1c} in patients with type 2 diabetes, further research has to be conducted with a larger sample size.

Conclusions

EasyNutrition was developed to introduce and evaluate the novel concept of nutrient profiling in the domain of dietary management apps and investigate its effect on the EH of people with diabetes by focusing their attention toward nutrition and, hence, helping them make better, healthier food choices. In addition, it was developed with cultural differences in mind, that is, all food recipes suggested by *EasyNutrition* are tailored to users' food preferences. Using *EasyNutrition*, the user can find web-based food recipes tailored to his/her food preferences; learn about their nutrition in a simple, easy-to-understand manner; and plan their weekly meals. *EasyNutrition* has been evaluated in a rigorous 3-stage step-up plan.

In the third stage, participants in the intervention group exhibited significant changes in their HbA_{1c}. Participants in the control group, however, exhibited minimal and nonstatistically significant changes in their HbA_{1c}. In addition, postresults of the independent samples *t* test showed a statistically significant difference between the control and intervention groups in terms of their HbA_{1c}. In conclusion, the results from this design research study illustrate the potential efficacy of using a customized nutrient profiling–inspired dietary app to focus individuals' attention toward nutrition and improve their EH.

Acknowledgments

This Project was funded by the Deanship of Scientific Research (DSR) at King Abdelaziz University, Jeddah, under grant no G-276-612-1440. The authors, therefore, acknowledge with thanks DSR for technical and financial support. Samir Chatterjee was partly funded for a portion of this study by a Fellowship from Schoeller Research Center for Business and Society in Nuremberg, Germany.

Conflicts of Interest

None declared.

References

1. Bader A, Gougeon R, Joseph L, Da Costa D, Dasgupta K. Nutritional education through internet-delivered menu plans among adults with type 2 diabetes mellitus: pilot study. JMIR Res Protoc 2013 Oct 11;2(2):e41 [FREE Full text] [doi: 10.2196/resprot.2525] [Medline: 24185033]

2. Evert AB, Boucher JL, Cypress M, Dunbar SA, Franz MJ, Mayer-Davis EJ, et al. Nutrition therapy recommendations for the management of adults with diabetes. Diabetes Care 2014 Jan;37(Suppl 1):S120-S143. [doi: 10.2337/dc14-S120] [Medline: 24357208]
3. Ogden CL, Carroll MD, Fryar CD, Flegal KM. Centers for Disease Control and Prevention. 2015 Nov. Prevalence of Obesity Among Adults and Youth: United States, 2011–2014 URL: https://www.cdc.gov/nchs/data/databriefs/db219.pdf [accessed 2016-12-13]
4. Kochanek KD, Murphy SL, Xu J, Tejada-Vera B. Centers for Disease Control and Prevention. 2016 Jun 30. National Vital Statistics Reports URL: https://www.cdc.gov/nchs/data/nvsr/nvsr65/nvsr65_04.pdf [accessed 2020-03-11]
5. American Diabetes Association. 2016. Statistics About Diabetes URL: http://www.diabetes.org/diabetes-basics/statistics/ [accessed 2016-12-12]
6. Arens-Volland AG, Spassova L, Bohn T. Promising approaches of computer-supported dietary assessment and management-Current research status and available applications. Int J Med Inform 2015 Dec;84(12):997-1008. [doi: 10.1016/j.ijmedinf.2015.08.006] [Medline: 26321486]
7. Epstein LH, Wing RR, Koeske R, Andrasik F, Ossip DJ. Child and parent weight loss in family-based behavior modification programs. J Consult Clin Psychol 1981 Oct;49(5):674-685. [doi: 10.1037//0022-006x.49.5.674] [Medline: 7287977]
8. Hevner A, Chatterjee S. Design Research in Information Systems. Boston, MA: Springer US; 2010.
9. US Department of Agriculture. USDA Food Composition Database. URL: https://ndb.nal.usda.gov/ndb/search/list [accessed 2016-11-03]
10. Woo YJ, Lee HS, Kim WY. Individual diabetes nutrition education can help management for type II diabetes. J Nutr Health 2006;39(7):641-648 [FREE Full text]
11. Alrige MA, Chatterjee S, Medina E, Nuval J. Applying the concept of nutrient-profiling to promote healthy eating and raise individuals' awareness of the nutritional quality of their food. AMIA Annu Symp Proc 2017;2017:393-402 [FREE Full text] [Medline: 29854103]
12. Årsand E, Frøisland DH, Skrøvseth SO, Chomutare T, Tatara N, Hartvigsen G, et al. Mobile health applications to assist patients with diabetes: lessons learned and design implications. J Diabetes Sci Technol 2012 Sep 1;6(5):1197-1206 [FREE Full text] [doi: 10.1177/193229681200600525] [Medline: 23063047]
13. Epstein LH, Myers MD, Raynor HA, Saelens BE. Treatment of pediatric obesity. Pediatrics 1998 Mar;101(3 Pt 2):554-570. [Medline: 12224662]
14. Epstein LH, Valoski AM, Vara LS, McCurley J, Wisniewski L, Kalarchian MA, et al. Effects of decreasing sedentary behavior and increasing activity on weight change in obese children. Health Psychol 1995 Mar;14(2):109-115. [doi: 10.1037//0278-6133.14.2.109] [Medline: 7789345]
15. Epstein LH, Wing RR, Koeske R, Ossip D, Beck S. A comparison of lifestyle change and programmed aerobic exercise on weight and fitness changes in obese children. Behav Ther 1982;13(5):651-665. [doi: 10.1016/s0005-7894(82)80022-1]
16. Hevner AR, March ST, Park J, Ram S. Design science in information systems research. Manag Inf Syst Q 2004 Mar;28(1):75-105. [doi: 10.2307/25148625]
17. Harris JA, Benedict FG. A biometric study of human basal metabolism. Proc Natl Acad Sci USA 1918 Dec;4(12):370-373 [FREE Full text] [doi: 10.1073/pnas.4.12.370] [Medline: 16576330]
18. Alrige M, Chatterjee S. Easy Nutrition: A Customized Dietary App to Highlight the Food Nutritional Value. In: International Conference on Design Science Research in Information Systems and Technology. 2018 Presented at: DESRIST'18; Chennai, India; June 3-6, 2018 p. 132-145.
19. Harris AD, McGregor JC, Perencevich EN, Furuno JP, Zhu J, Peterson DE, et al. The use and interpretation of quasi-experimental studies in medical informatics. J Am Med Inform Assoc 2006;13(1):16-23 [FREE Full text] [doi: 10.1197/jamia.M1749] [Medline: 16221933]
20. Walker S, Sechrist K, Pender N. The Health-Promoting Lifestyle Profile. Nursing Research 1987;36(2):76-81. [doi: 10.1097/00006199-198703000-00002] [Medline: 6392204]
21. AbuSabha R, Achterberg C. Review of self-efficacy and locus of control for nutrition- and health-related behavior. J Am Diet Assoc 1997 Oct;97(10):1122-1132. [doi: 10.1016/S0002-8223(97)00273-3] [Medline: 9336559]
22. Rosenstock IM, Strecher VJ, Becker MH. Social learning theory and the Health Belief Model. Health Educ Q 1988;15(2):175-183. [doi: 10.1177/109019818801500203] [Medline: 3378902]

Abbreviations

ADA: American Diabetes Association
API: Application Programming Interface
BMR: basal metabolic rate
CVD: cardiovascular disease
DSR: design science research
DTC: diabetes treatment center
EH: eating habits

HbA_{1c}: glycated hemoglobin
HPLP: Health-Promoting Lifestyle Profile
IRB: Institutional Review Board
NSE: nutrition self-efficacy
SOC: standard of care
T2DM: type 2 diabetes mellitus

Edited by G Eysenbach; submitted 16.07.19; peer-reviewed by D Bitar, A Kulanthaivel, M Lee; comments to author 21.10.19; revised version received 15.12.19; accepted 26.01.20; published 19.05.20.

Please cite as:
Alrige M, Alharbey R, Chatterjee S
The Effect of a Customized Nutrient-Profiling Approach on the Glycated Hemoglobin Levels of Patients With Type 2 Diabetes: Quasi-Experimental Study
J Med Internet Res 2020;22(5):e15497
URL: https://www.jmir.org/2020/5/e15497
doi:10.2196/15497
PMID:32427107

Original Paper

Online Impact and Presence of a Specialized Social Media Team for the Journal of Neurosurgery: Descriptive Analysis

Joseph R Linzey[1], MD, MS; Faith Robertson[2], MD, MSc; Ali S Haider[3], BS; Christopher Salvatore Graffeo[4], MD, MS; Justin Z Wang[5], MD; Gillian Shasby[6], BA; Naif M Alotaibi[2], MD, MSc; Aaron A Cohen-Gadol[7], MD, MSc, MBA; James T Rutka[5], MD, PhD

[1]Department of Neurosurgery, University of Michigan, Ann Arbor, MI, United States
[2]Department of Neurosurgery, Massachusetts General Hospital and Harvard Medical School, Boston, MA, United States
[3]Department of Industrial and Systems Engineering, Texas A&M University, College Station, TX, United States
[4]Department of Neurologic Surgery, Mayo Clinic, Rochester, MN, United States
[5]Division of Neurosurgery, Faculty of Kinesiology and Physical Education, University of Toronto, Toronto, ON, Canada
[6]Journal of Neurosurgery Publishing Group, Charlottesville, VA, United States
[7]Department of Neurological Surgery, Indiana University, Indianapolis, IN, United States

Corresponding Author:
Naif M Alotaibi, MD, MSc
Department of Neurosurgery
Massachusetts General Hospital and Harvard Medical School
55 Fruit St
Boston, MA, 02114
United States
Phone: 1 8579489955
Email: nalotaibi@mgh.harvard.edu

Abstract

Background: Social media use continues to gain momentum in academic neurosurgery. To increase journal impact and broaden engagement, many scholarly publications have turned to social media to disseminate research. The Journal of Neurosurgery Publishing Group (JNSPG) established a dedicated, specialized social media team (SMT) in November 2016 to provide targeted improvement in digital outreach.

Objective: The goal of this study was to examine the impact of the JNSPG SMT as measured by increased engagement.

Methods: We analyzed various metrics, including impressions, engagements, retweets, likes, profile clicks, and URL clicks, from consecutive social media posts from the JNSPG's Twitter and Facebook platforms between February 1, 2015 and February 28, 2019. Standard descriptive statistics were utilized.

Results: Between February 2015 and October 2016, when a specialized SMT was created, 170 tweets (8.1 tweets/month) were posted compared to 3220 tweets (115.0 tweets/month) between November 2016 and February 2019. All metrics significantly increased, including the impressions per tweet (mean 1646.3, SD 934.9 vs mean 4605.6, SD 65,546.5; P=.01), engagements per tweet (mean 35.2, SD 40.6 vs mean 198.2, SD 1037.2; P<.001), retweets (mean 2.5, SD 2.8 vs mean 10.5, SD 15.3; P<.001), likes (mean 2.5, SD 4.0 vs mean 18.0, SD 37.9; P<.001), profile clicks (mean 1.5, SD 2.0 vs mean 5.2, SD 43.3; P<.001), and URL clicks (mean 13.1, SD 14.9 vs mean 38.3, SD 67.9; P<.001). Tweets that were posted on the weekend compared to weekdays had significantly more retweets (mean 9.2, SD 9.8 vs mean 13.4, SD 25.6; P<.001), likes (mean 15.3, SD 17.9 vs mean 23.7, SD 70.4; P=.001), and URL clicks (mean 33.4, SD 40.5 vs mean 49.5, SD 117.3; P<.001). Between November 2015 and October 2016, 49 Facebook posts (2.3 posts/month) were sent compared to 2282 posts (81.5 posts/month) sent between November 2016 and February 2019. All Facebook metrics significantly increased, including impressions (mean 5475.9, SD 5483.0 vs mean 8506.1, SD 13,113.9; P<.001), engagements (mean 119.3, SD 194.8 vs mean 283.8, SD 733.8; P<.001), and reach (mean 2266.6, SD 2388.3 vs mean 5344.1, SD 8399.2; P<.001). Weekend Facebook posts had significantly more impressions per post (mean 7967.9, SD 9901.0 vs mean 9737.8, SD 19,013.4; P=.03) and a higher total reach (mean 4975.8, SD 6309.8 vs mean 6108.2, SD 12,219.7; P=.03) than weekday posts.

Conclusions: Social media has been established as a crucial tool for the propagation of neurosurgical research and education. Implementation of the JNSPG specialized SMT had a demonstrable impact on increasing the online visibility of social media content.

(J Med Internet Res 2020;22(5):e17741) doi:10.2196/17741

KEYWORDS

social media; Twitter; Facebook; research dissemination

Introduction

Social media use continues to gain momentum in both the general public and in academic medicine. When the Pew Research Center began tracking social media statistics in 2005, 5% of Americans were technology adopters, whereas 72% of the public currently uses some variation of social media (eg, Facebook, Twitter, Youtube, LinkedIn, Instagram) [1]. These digital applications are multidimensional tools that—in addition to aiding personal relationships and providing entertainment—can be leveraged to disseminate research to both medical professionals and patient populations, strengthen professional networks, and promote increased connectivity within a professional field. The use of social media to improve literature dissemination, foster a digital presence, and enhance programs' rankings has been especially prominent for surgical specialties, with many recent publications and digital topic trends in general [2,3], cardiothoracic [4], urology [5], and plastic surgery [6], as well as other subspecialties.

For neurosurgery specifically, there is particular interest in engaging younger surgeons, publicizing literature, and enhancing more global collaborations. The 2018 article "Millennials in neurosurgery: Is there hope?" [7] expressed concern for the traits and opinions of this new generation, but millennials were commended for their increased use and familiarity with communications, media, and technology along with an expectation to use these tools daily to make their work more efficient. Importantly, as social media tools are embraced by all demographics, and the highest adopters are young professionals (88% of 18- to 29-year-olds), social media provides an opportunity to further engage and train the new generation. Furthermore, we are experiencing a shift in the paradigm for evaluating the impact of research. Academic science and biomedicine publications have been historically ranked by citation-based metrics (eg, H-index, impact score); however, altmetrics are gaining prevalence within medicine [8,9]. Altmetrics are an alternative set of quantitative and qualitative scores such as attention, mentions, and retweets that assess public engagement. In 2017, Wang et al [10] conducted a qualitative analysis of the highest trending works in neurosurgery along with a correlation analysis with their social media metrics. In considering altmetric scores, they found an average score of 4.7 (SD 22.4); journals with a social media account had significantly higher altmetric scores for their articles than journals without such an account (P<.001). The top 100 neurosurgical articles in altmetrics belonged primarily to those with active accounts, with *Journal of Neurosurgery* ranking the highest (33%), followed by *Neurosurgery* (29%) [10]. ResearchGate is another digital platform that is widely used to share articles, increase neurosurgical networking, and facilitate research collaboration [9]. These digital partnerships are continuing to grow and influence how we conduct research.

Overall, social media presence for academic journals is important for greater outreach and engagement. The *Journal of Neurosurgery* Publishing Group (JNSPG) established a dedicated social media team (SMT) to provide targeted improvement of digital outreach. The goal of the present study was to examine the impact of establishing a dedicated SMT and elucidate the characteristics of Twitter and Facebook posts that optimize audience engagement. By providing a clearer understanding of how post timing, images, hashtags, and more influence interaction, these data can inform educational and business strategies for social media influencers and other academic journals.

Methods

The Social Media Team

The JNSPG SMT was created on November 1, 2016 and is comprised of editors who examine both new publications and previously published articles in the core and subtopic journals of the JNSPG. There have been 4-6 social media editors, all of whom are either neurosurgical residents or senior medical students committed to a career in neurosurgery. They identified interesting articles and appealing or representative figures and images, and created a brief headline or summary of the publication with hashtags corresponding to the article type. Each editor was responsible for creating content for 1-3 hashtags for the upcoming week. Post content was checked and verified by a social media manager who was also a neurosurgical resident. The posts were scheduled to be published via Hootsuite software (Hootsuite Inc, Vancouver, BC, Canada) [11] at various times throughout the day. Editors were also responsible for generating visual abstracts on behalf of the journal and microblogging using their personal accounts to comment on occasional articles. Visual abstracts underwent an extra layer of scrutiny before publication. The journal's editorial and staff teams analyzed the visual abstracts to ensure congruency with the manuscript's original content and appeal of the visual components of the abstract.

Study Design

We analyzed metrics from consecutive social media posts from the JNSPG's Twitter and Facebook platforms between February 1, 2015 and February 28, 2019. All data were obtained from the analytics collected by Twitter and Facebook. Since all data used were publicly accessible, Institutional Review Board approval was not needed. Additionally, since no patient information was accessed, no patient consent was sought.

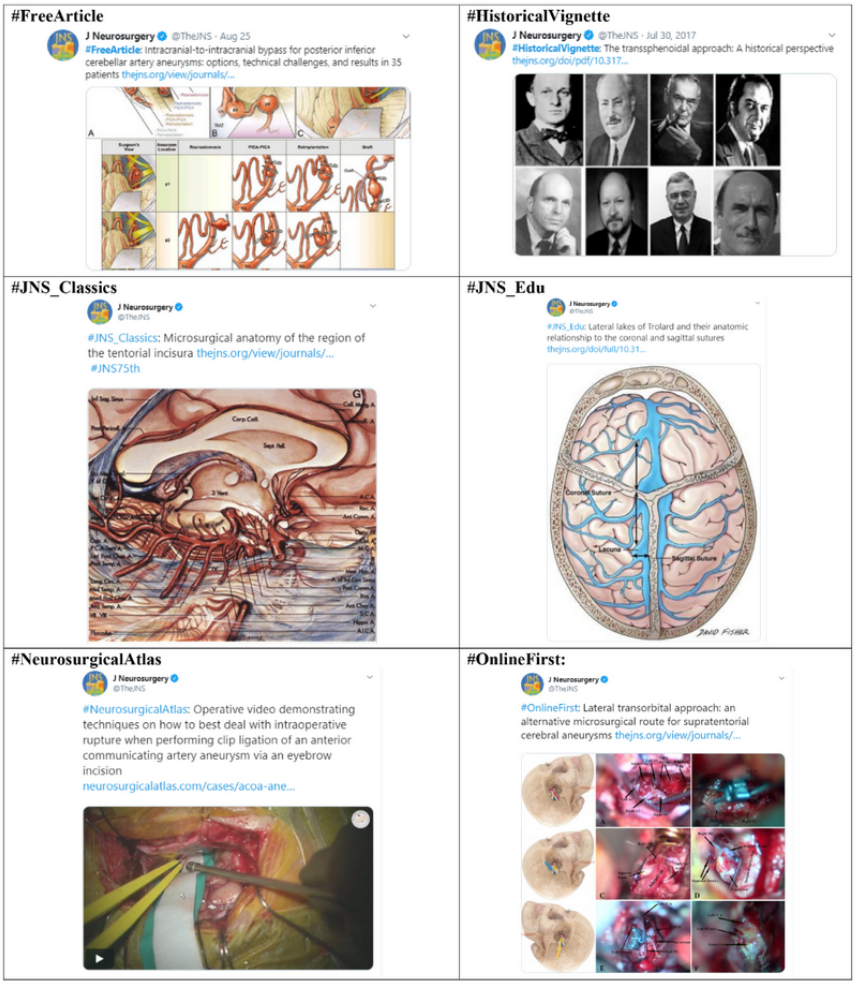
Social Media Data

Twitter

The variables collected describing the Twitter posts included the content of the tweet, the date and time the tweet was posted, and the number of impressions (number of times a post was viewed by a user, whether the post was clicked on or not), engagements (number of times a post was clicked on to magnify the image or text or view a video), retweets, likes, profile clicks, and URL clicks each post received. The definitions for each metric were obtained from the Twitter analytics data we received during data abstraction.

The text of each tweet was searched for hashtags and divided into 14 different categories: #FreeArticle, #HistoricalVignette, #JNS_Classics, #JNS_Edu, #NeurosurgicalAtlas, #OnlineFirst, #OperativeVideo, #VideoAbstract, #VisualAbstract, #JNS_History, #GoogleAlerts, #JNS75th, #NeurosurgicalFocus, and other (see Figure 1 and Multimedia Appendix 1). The date the content was posted was dichotomized into weekdays (Monday through Friday) and weekends (Saturday and Sunday). The time the tweet was posted was also split into content sent during the workday (7 am to 5 pm) and after the workday (5:01 pm to 6:59 am) in Eastern Standard Time/Eastern Daylight Savings Time.

Figure 1. Examples of different types of Journal of Neurosurgery Publishing Group (JNSPG) Twitter posts.



Facebook

The variables collected describing the Facebook posts included the content of the post, date and time the content was posted, number of impressions (number of times a post was viewed by a user, whether the post was clicked on or not), number of engagements (number of times a post was clicked on to magnify the image or text or view a video), and total number of unique people who received each post. The definitions for each metric were obtained from the Facebook analytics data we received during data abstraction

The text of each Facebook post was searched for hashtags and divided into 13 different categories: #FreeArticle, #HistoricalVignette, #JNS_Classics, #JNS_Edu, #NeurosurgicalAtlas, #OnlineFirst, #OperativeVideo, #VideoAbstract, #VisualAbstract, #JNS_History, #JNS75th, #NeurosurgicalFocus, and Other. The date the content was posted was dichotomized into weekdays (Monday through Friday) and weekends (Saturday and Sunday). The time the Facebook post was sent was also split into content sent during the workday (7 am to 5 pm) and after the workday (5:01 pm to 6:59 am) in Eastern Standard Time/Eastern Daylight Savings Time.

Statistical Analysis

Twitter and Facebook posts were dichotomized into pre- and postintervention groups using the date that the social media

team was initiated (November 1, 2016) as the intervention start time. Standard univariate descriptive statistical analyses, including *t* tests and one-way analysis of variance, were performed, and the data are summarized as means and SD. *P* values are based on two-sided tests, and values less than .05 were considered significant. All data were analyzed using SAS 9.4 software (SAS Institute Inc, Cary, NC, USA).

Results

Twitter

Between February 2015 and February 2019, a total of 3390 tweets were posted from the JNSPG Twitter account. In the preintervention period (February 2015 to October 2016), 170 tweets or 8.1 tweets per month were posted. After implementation of the SMT, 3220 tweets or 115.0 tweets per month were posted between November 2016 and February 2019.

The impressions per tweet significantly increased following creation of the SMT (Table 1), and interactions with Twitter content also increased significantly, including average engagements, retweets, and likes per tweet. Users were also significantly more likely to click on the JNSPG's Twitter profile after SMT implementation. Importantly, readers were significantly more likely to click on the URLs included in the tweet, which took the user to the original journal article.

Table 1. Twitter analytics before and after creation of a social media team.

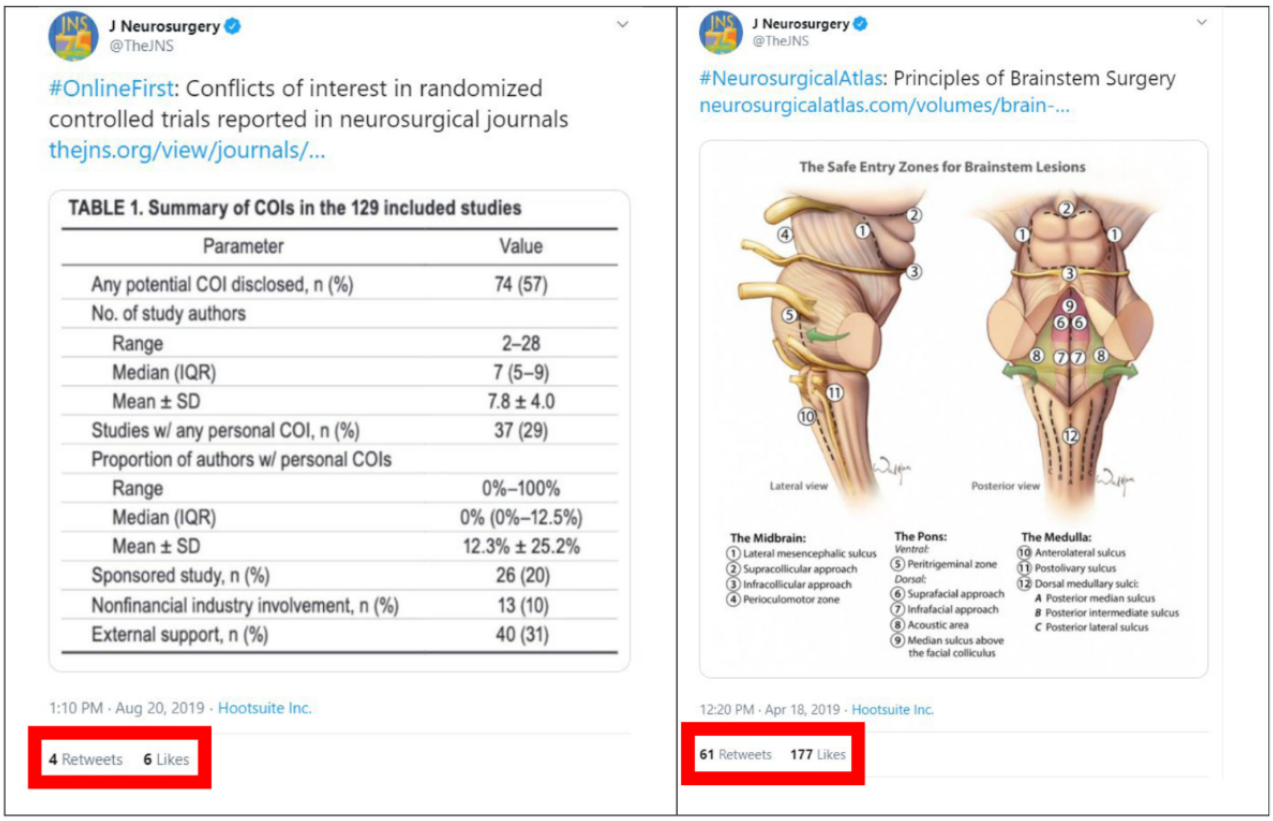
Metric	Before social media team, mean (SD)	After social media team, mean (SD)	*P* value
Impressions per tweet	1646.3 (934.9)	4605.6 (65546.5)	.01
Engagements per tweet	35.2 (40.6)	198.2 (1037.2)	<.001
Retweets per tweet	2.5 (2.8)	10.5 (15.3)	<.001
Likes per tweet	2.5 (4.0)	18.0 (37.9)	<.001
Profile clicks per tweet	1.5 (2.0)	5.2 (43.3)	<.001
URL clicks per tweet	13.1 (14.9)	38.3 (67.9)	<.001

Interactions between JNSPG tweets varied significantly depending on the content and hashtags used for different tweets. In almost all categories (engagements, retweets, likes, profile clicks, and URL clicks per tweet), tweets with the hashtags #JNS_Edu, #NeurosurgicalAtlas, #OperativeVideo, and #VisualAbstract were significantly more likely to receive high levels of interaction from users (*P*<.001, Table 2). Content in the #OnlineFirst, #GoogleAlert, #VideoAbstract, or Other categories was less likely to elicit similarly high levels of interactions (Figure 2).

Table 2. Twitter analytics, mean (SD) per tweet, by hashtag.

Hashtag	N	Impressions	Engagements	Retweets	Likes	Profile clicks	URL clicks
#FreeArticle	701	3750.6 (2892.6)	215.6 (231.2)	11.7 (9.1)	19.6 (16.7)	4.5 (5.9)	43.8 (37.9)
#Historical Vignette	60	5157.8 (4489.8)	234.5 (193.9)	14.7 (10.7)	24.5 (16.2)	5.0 (6.0)	30.0 (21.2)
#JNS_Classics	66	3518.0 (3123.7)	217.4 (385.8)	12.1 (12.0)	22.0 (25.7)	5.1 (9.5)	27.7 (29.5)
#JNS_Edu	15	6159.1 (2236.1)	503.7 (326.3)	24.1 (7.3)	39.7 (12.8)	9.9 (7.2)	65.7 (46.9)
#NeurosurgicalAtlas	220	6726.9 (4147.6)	447.7 (385.5)	23.6 (13.8)	48.4 (27.9)	11.5 (13.1)	83.6 (68.0)
#OnlineFirst	1173	3174.2 (2824.2)	169.3 (214.1)	9.5 (8.3)	15.1 (14.3)	3.9 (5.7)	32.5 (36.5)
#OperativeVideo	89	4832.3 (22936.4)	724.4 (6080.7)	14.3 (68.9)	34.2 (199.2)	29.8 (257.3)	55.6 (318.5)
#VideoAbstract	111	1863.1 (1105.7)	35.5 (31.1)	3.5 (1.9)	5.0 (3.6)	1.5 (1.8)	10.0 (8.2)
#Visual Abstract	12	8049.3 (3555.6)	347.8 (173.5)	25.8 (13.3)	32.3 (16.8)	10.1 (4.4)	43.5 (28.9)
Other	588	8873.8 (153016.6)	103.1 (171.5)	6.8 (9.9)	9.0 (13.2)	3.7 (6.9)	27.3 (39.7)
#JNSHistory	15	3611.3 (2406.4)	161.7 (146.0)	10.1 (7.3)	16.5 (15.2)	5.7 (4.9)	7.3 (10.5)
#GoogleAlerts	405	1620.8 (1848.0)	47.2 (70.3)	4.0 (5.5)	6.2 (8.4)	1.6 (3.7)	27.2 (41.1)
#JNS75th	15	5277.2 (2087.4)	255.2 (141.2)	16.0 (10.9)	32.1 (18.7)	10.7 (6.4)	23.3 (21.7)
#NeurosurgicalFocus	20	4435.7 (1860.7)	253.1 (194.3)	13.1 (9.2)	28.9 (18.7)	7.4 (5.2)	39.8 (28.6)
P value	.98	<.001	<.001	<.001	<.001	<.001	<.001

Figure 2. Differences in retweets and likes between #OnlineFirst and #NeurosurgicalAtlas.



Users were significantly more likely to retweet content that was posted from the JNSPG Twitter account after work hours (Table 3) compared to content posted during work hours. However, the average number of impressions, engagements, likes, profile clicks, or URL clicks did not significantly vary based on the time of day the tweet was posted. Tweets that were posted on the weekend compared to weekdays had significantly more retweets, likes, and URL clicks per tweet (Table 3). However, there was no significant difference in the average number of impressions, engagements, or profile clicks based on whether a tweet was posted on a weekday or during the weekend.

Table 3. Twitter analytics, mean (SD), as a function of time and day.

Analytic	Tweet sent during work hours (n=2024)	Tweet sent after work hours (n=1366)	*P* value	Weekday tweets (n=2626)	Weekend tweets (n=764)	*P* value
Impressions per tweet	3186.3 (3107.6)	6340.3 (10,0562.0)	.25	4517.8 (72,448.4)	4248.8 (8368.4)	.85
Engagements per tweet	165.6 (234.0)	226.3 (1567.5)	.16	157.8 (232.0)	301.1 (2084.2)	.06
Retweets per tweet	9.3 (9.7)	11.4 (20.4)	<.001	9.2 (9.8)	13.4 (25.6)	<.001
Likes per tweet	16.0 (18.1)	19.0 (54.1)	.05	15.3 (17.9)	23.7 (70.4)	.001
Profile clicks per tweet	3.9 (6.3)	6.6 (66.1)	.14	4.0 (7.0)	8.5 (87.9)	.16
URL clicks per tweet	35.5 (42.5)	39.3 (91.0)	.15	33.4 (40.5)	49.5 (117.3)	<.001

Facebook

Between November 2015 and February 2019, a total of 2331 Facebook posts were sent from the JNSPG Facebook account. Between November 2015 and October 2016, when the specialized SMT was created, 49 posts or 2.3 posts per month were sent. Between November 2016 and February 2019, after implementation of the SMT, 2282 posts or 81.5 posts per month were sent.

After implementation of the dedicated SMT, average impressions for Facebook posts, average engagements, and average reach per post increased significantly (Table 4).

Mirroring the trend found in the Twitter data, users were most likely to engage with posts with the hashtags #JNS_Edu, #NeurosurgicalAtlas, and #OperativeVideo compared to hashtags such as #OnlineFirst, #VideoAbstract, and #JNS_History (*P*<.001, Table 5). These same post types had significantly more impressions and a greater reach among Facebook users (Figure 2).

There were no significant differences in the number of impressions, engagements, or total reach of Facebook posts based on whether the post was sent during work hours or after work hours (Table 6). However, Facebook posts sent on the weekend compared to weekdays were significantly more likely to have a higher average number of impressions per post and higher total reach. The average number of engagements per post did not vary between weekend and weekday posts (Table 6).

Table 4. Facebook analytics, mean (SD), before and after creation of a social media team.

Analytic	Before social media team (n=49)	After social media team (n=2282)	*P* value
Impressions per post	5475.9 (5483.0)	8506.1 (13,113.9)	<.001
Engagements per post	119.3 (194.8)	283.8 (733.8)	<.001
Total reach per post	2266.6 (2388.3)	5344.1 (8399.2)	<.001

Table 5. Facebook post analytics, mean (SD), by hashtag.

Hashtag	N	Impressions	Engagements	Total reach
#FreeArticle	670	9088.3 (12012.8)	297.8 (501.6)	5644.3 (7379.7)
#Historical Vignette	57	7769.6 (6208.1)	203.1 (195.1)	229.1 (389.7)
#JNS_Classics	62	5644.3 (7379.7)	4862.3 (3600.6)	4849.3 (4894.9)
#JNS_Edu	17	21131.5 (12633.4)	735.7 (529.5)	12474.6 (7441.6)
#NeurosurgicalAtlas	220	11997.4 (8087.2)	442.0 (356.6)	7718.1 (4896.9)
#OnlineFirst	960	6677.1 (6365.1)	196.1 (259.5)	4197.7 (3925.4)
#OperativeVideo	83	13890.4 (40761.4)	653.4 (2949.7)	8787.3 (26849.9)
#VideoAbstract	8	4468.6 (2696.0)	141.4 (109.6)	2863.9 (1829.5)
#Visual Abstract	7	8796.0 (7721.9)	228.0 (296.5)	5402.0 (4650.2)
Other	208	8360.3 (20484.6)	303.3 (1005.6)	4964.1 (13524.1)
#JNSHistory	11	4046.7 (2941.9)	128.8 (95.3)	2629.2 (1957.6)
#JNS75th	8	8146.1 (4647.8)	294.6 (241.4)	5462.6 (3079.1)
#NeurosurgicalFocus	20	8428.2 (6108)	319.5 (329.0)	5592.9 (3881.2)
P value	N/A[a]	<.001	<.001	<.001

[a]N/A: not applicable.

Table 6. Facebook post analytics, mean (SD), as a function of time and day.

Analytic	Posts sent during work hours (n=1700)	Posts sent after work hours (n=631)	*P* value	Weekday posts (n=1706)	Weekend posts (n=625)	*P* value
Impressions per post	8533.0 (12,758.6)	8198.3 (13,659.0)	.59	7967.9 (9901.0)	9737.8 (19,013.4)	.03
Engagements per post	284.3 (749.5)	269.8 (663.1)	.65	259.2 (456.0)	338.1 (1183.6)	.10
Total reach per post	5330.5 (8107.4)	5142.0 (8904.7)	.64	4975.8 (6309.8)	6108.2 (12,219.7)	.03

Discussion

Principal Findings

Social media, especially Twitter and Facebook, are increasingly utilized for the dissemination of neurosurgical research. There are many potential strategies for journals to increase their social media presence, including frequent postings, the creation of a general marketing team, and the creation of a specialized SMT. Our study illustrates the success of implementing a specialized SMT for the JNSPG. All key outcomes (the quantity of content published, impressions, likes, retweets, engagements, profile clicks, URL clicks, and total reach) were significantly increased after development of a dedicated, specialized SMT.

Social media use among neurosurgeons, neurosurgical departments, and neurosurgical journals has increased exponentially over the past few years [12-17]. With the rise in social media use among neurosurgeons, there is an increased need for neurosurgical journals to produce regular, high-quality content. Altmetric [18] has been established as a grading mechanism for determining article impact on social media platforms [19]. It has been noted that articles published in journals with a robust social media presence had higher altmetric scores, further demonstrating the meaningful role of a specialized SMT in propagating an article's online impact [10]. In addition, multiple studies have reported a correlation between an active social media presence and increased research productivity among academic medical institutions [12,13,20,21].

This provides evidence that traditional bibliometrics also benefit when an article has an impact on social media, with the presumption that the article is read, downloaded, and engaged with more due to its greater online reach.

Given the complex and unique knowledge base from which neurosurgery draws, it would be difficult for a general marketing team to develop the expertise necessary to produce content that engages and challenges the neurosurgical community. There is a distinct need for a specialized group of neurosurgical students, trainees, and attending surgeons to help create and moderate the content on the social media platforms of neurosurgical journals. Given the large amount of time needed to manage a major social media presence on top of a busy clinical practice, it is necessary to maintain a fairly large team of volunteers to appropriately divide the workload. For many trainees, the ability to "get their foot in the door" and work directly with a specialized journal in their field is sufficient motivation to provide the enthusiasm needed for sustained teams with necessary longevity. After the JNSPG created their SMT, there was a significant increase in all metrics that indicate greater engagement with the posts.

There was a significant difference in the engagement level based on the different types of posts and hashtags used. Universally, posts with impressive illustrations or videos (#JNS_Edu, #NeurosurgicalAtlas, #OperativeVideo) were significantly more likely to produce high engagement levels compared to other, less visually appealing posts (#OnlineFirst, #GoogleAlert, or Other). Having a specialized, dedicated team that is focused on optimizing the visual presentation of social media content and knowing what specific types of content are most likely to resonate with the neurosurgical community is imperative. Additionally, if feasible, it may benefit journals and institutions to have staff dedicated to medical illustrations.

Overall, there was no significant difference in the average number of impressions, likes, profile clicks, URL clicks, or total reach for Twitter or Facebook content based on whether the post was sent during typical work hours or after hours. There were significantly more retweets for after-hour posts. However, there were significantly more retweets, likes, and URL clicks for Twitter posts, along with more impressions and a higher total reach for Facebook posts if they were posted on the weekend compared to a weekday. While the time of day may not affect the likelihood of a surgeon perusing social media, these results indicate that neurosurgeons are able to take more time to not only quickly like or retweet interesting content on the weekend, but to also spend additional time engaging with the original research articles by clicking on the URLs embedded in the tweets. The time of day that content is posted may not significantly affect neurosurgeons' engagement patterns because their busy operative schedules may allow only sporadic scrolling through their feeds at random times of the day. However, on weekends, many surgeons have extra time to more fully engage with interesting manuscripts and images that are produced for social media. Again, a dedicated SMT can take advantage of this knowledge to target especially high-impact content to the weekend when a neurosurgeon is most likely to have sufficient time to engage with the social media content.

As online resources and social media continue to play a greater role in research dissemination and clinical trainee education, it is important to emphasize the need for free educational sources that can be easily distributed via social media channels. Online educational resources such as The Neurosurgical Atlas have partnered with neurosurgical journals, including *Journal of Neurosurgery* and *Operative Neurosurgery*, to create an interactive and comprehensive learning tool for distribution on social media [22]. This educational collaboration has provided access to large amounts of instructive material to learners anywhere in the world without any direct costs. Specialized teams of social media editors can sift through and collate the vast amounts of educational data for optimal distribution.

As the *Journal of Neurosurgery* SMT continues to grow and develop, further research initiatives will be pursued. The benefits of visual abstracts, specifically in manuscripts without compelling visual figures, will be examined. Further research is needed to determine the benefit of operative videos compared to static operative or illustrated images. Finally, additional work will be pursued to delineate the relationship between the level of engagement and the time of day or day of the week when content is posted taking the worldwide viewership into account.

Limitations

The main limitation of this study is that the data are from a single, neurosurgical journal's experience. In the future, additional studies that compare and contrast the experiences of multiple journals with and without SMTs would strengthen the results presented in this study. Additionally, this study analyzed the JNSPG's experience with only Twitter and Facebook. Many other journals and neurosurgical departments utilize Instagram and other social media platforms that are not reflected in these analyses. Furthermore, there is currently no strong, accepted method to correlate increased social media visibility with traditional bibliometric data. Altmetrics utilize social media engagement levels to create a score, but it is unclear if higher altmetric scores correlate with more citations in neurosurgical research. We found an association between post timing and post engagement; however, we do not have any data on when the posts were actually viewed by the user. This association is further hindered by the large international following of the JNSPG's social media platforms, which introduces variability with different time zones. Additionally, we do not have any data on the specific metrics resulting from individual members of the SMT using their personal accounts to comment or boost the visibility of specific posts. Given the relative lack of followers that individual members of the SMT have compared to the *Journal of Neurosurgery* account, we are relatively confident that these microblogs did not significantly skew the data.

Conclusions

Social media is established as a crucial tool for the propagation of neurosurgical research and education. Implementation of a specialized SMT has a demonstrable impact on increasing the online visibility of social media content.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional examples of the 14 categories of Journal of Neurosurgery Publishing Group (JNSPG) Twitter posts.
[PDF File (Adobe PDF File), 293 KB - jmir_v22i5e17741_app1.pdf]

References

1. Pew Research Center. 2019 Jun 12. Social media use in 2018 URL: https://www.pewinternet.org/fact-sheet/social-media/ [accessed 2019-10-01]
2. Logghe HJ, Boeck MA, Gusani NJ, Hardaway JC, Hughes KA, Mouawad NJ, et al. Best Practices for Surgeons' Social Media Use: Statement of the Resident and Associate Society of the American College of Surgeons. J Am Coll Surg 2018 Mar;226(3):317-327. [doi: 10.1016/j.jamcollsurg.2017.11.022] [Medline: 29224795]
3. Chiang AL, Galler Rabinowitz L, Kumar A, Chan WW. Association Between Institutional Social Media Involvement and Gastroenterology Divisional Rankings: Cohort Study. J Med Internet Res 2019 Sep 06;21(9):e13345 [FREE Full text] [doi: 10.2196/13345] [Medline: 31493321]
4. Luc JGY, Archer MA, Arora RC, Bender EM, Blitz A, Cooke DT, et al. Social Media Improves Cardiothoracic Surgery Literature Dissemination: Results of a Randomized Trial. Ann Thorac Surg 2019 Aug 09:589-595. [doi: 10.1016/j.athoracsur.2019.06.062] [Medline: 31404547]
5. Teoh JY, Mackenzie G, Tortolero L, Rivas JG. Social Media Analytics: What You Need to Know as a Urologist. Eur Urol Focus 2019 Aug 28:Epub. [doi: 10.1016/j.euf.2019.08.005] [Medline: 31471218]
6. Chen AD, Ruan QZ, Bucknor A, Chattha AS, Bletsis PP, Furnas HJ, et al. Social Media: Is the Message Reaching the Plastic Surgery Audience? Plast Reconstr Surg 2019 Sep;144(3):773-781. [doi: 10.1097/PRS.0000000000005988] [Medline: 31461046]
7. Spiotta AM, Kalhorn S, Patel S. Millenials in Neurosurgery: Is there Hope? Neurosurgery 2018 Aug 01;83(2):E71-E73. [doi: 10.1093/neuros/nyy228] [Medline: 29800465]
8. Thelwall M, Haustein S, Larivière V, Sugimoto CR. Do altmetrics work? Twitter and ten other social web services. PLoS One 2013;8(5):e64841 [FREE Full text] [doi: 10.1371/journal.pone.0064841] [Medline: 23724101]
9. Joshi ND, Lieber B, Wong K, Al-Alam E, Agarwal N, Diaz V. Social Media in Neurosurgery: Using ResearchGate. World Neurosurg 2019 Jul;127:e950-e956. [doi: 10.1016/j.wneu.2019.04.007] [Medline: 30965167]
10. Wang J, Alotaibi NM, Ibrahim GM, Kulkarni AV, Lozano AM. The Spectrum of Altmetrics in Neurosurgery: The Top 100. World Neurosurg 2017 Jul;103:883-895. [doi: 10.1016/j.wneu.2017.04.157] [Medline: 28478251]
11. Houtsuite. URL: https://hootsuite.com/ [accessed 2020-04-15]
12. Alotaibi NM, Badhiwala JH, Nassiri F, Guha D, Ibrahim GM, Shamji MF, et al. The Current Use of Social Media in Neurosurgery. World Neurosurg 2016 Apr;88:619-624. [doi: 10.1016/j.wneu.2015.11.011] [Medline: 26585734]
13. Alotaibi NM, Samuel N, Guha D, Nassiri F, Badhiwala JH, Tam J, et al. Social Media for Academic Neurosurgical Programs: The University of Toronto Experience. World Neurosurg 2016 Sep;93:449-457. [doi: 10.1016/j.wneu.2016.06.134] [Medline: 27423202]
14. Ban VS, Lega B, Batjer HH. Maximizing the Potential of Social Media and Social Networks in Neurosurgery. World Neurosurg 2016 Jul;91:609-610. [doi: 10.1016/j.wneu.2016.02.117] [Medline: 26968446]
15. Fargen KM, Wilson TA, de Leacy R, Jadhav AP, Ducruet AF. Social media and expanding the digital clique. J Neurointerv Surg 2017 Oct;9(10):913-914. [doi: 10.1136/neurintsurg-2017-013421] [Medline: 28899931]
16. Fargen KM, Ducruet AF, Hyer M, Hirsch JA, Tarr RW. Expanding the social media presence of the Journal of Neurointerventional Surgery: editor's report. J Neurointerv Surg 2017 Feb;9(2):215-218. [doi: 10.1136/neurintsurg-2015-012251] [Medline: 26927814]
17. Zusman EE. Social Media in Academic Neurosurgery. World Neurosurg 2016 Jul;91:606-608. [doi: 10.1016/j.wneu.2016.02.118] [Medline: 26996732]
18. Altmetric. Who’s talking about your research? URL: https://www.altmetric.com/ [accessed 2020-08-01]
19. Melero R. Altmetrics - a complement to conventional metrics. Biochem Med (Zagreb) 2015;25(2):152-160 [FREE Full text] [doi: 10.11613/BM.2015.016] [Medline: 26110028]
20. Alotaibi NM, Guha D, Fallah A, Aldakkan A, Nassiri F, Badhiwala JH, et al. Social Media Metrics and Bibliometric Profiles of Neurosurgical Departments and Journals: Is There a Relationship? World Neurosurg 2016 Jun;90:574-579. [doi: 10.1016/j.wneu.2016.01.087] [Medline: 26855310]
21. Bean JR. Academic Output and Social Media: A Marriage of Opposites. World Neurosurg 2016 Jun;90:651-653. [doi: 10.1016/j.wneu.2016.02.004] [Medline: 26862024]

22. Linzey JR, Graffeo CS, Wang JZ, Haider AS, Alotaibi NM. Neurosurgery and the rise of academic social media: what neurosurgeons should know. J Neurosurg 2018 Oct;129(4):1-5. [doi: 10.3171/2018.2.JNS172817] [Medline: 29979122]

Abbreviations

JNSPG: Journal of Neurosurgery Publishing Group
SMT: social media team

Edited by G Eysenbach; submitted 10.01.20; peer-reviewed by E Da Silva, S Manca; comments to author 22.02.20; revised version received 03.03.20; accepted 05.03.20; published 19.05.20.

Please cite as:
Linzey JR, Robertson F, Haider AS, Graffeo CS, Wang JZ, Shasby G, Alotaibi NM, Cohen-Gadol AA, Rutka JT
Online Impact and Presence of a Specialized Social Media Team for the Journal of Neurosurgery: Descriptive Analysis
J Med Internet Res 2020;22(5):e17741
URL: http://www.jmir.org/2020/5/e17741/
doi:10.2196/17741
PMID:32163371

Original Paper

Internet of Things Smart Sunscreen Station: Descriptive Proof-of-Concept Study

Helen Ford[1*], MSc; Jeremy Herbert[2*], PhD; Caitlin Horsham[1*], MSc; Alexander Wall[2], GradDip; Elke Hacker[1], PhD

[1]Institute of Health and Biomedical Innovation, School of Public Health and Social Work, Queensland University of Technology, Brisbane, Australia
[2]Designworks Group Pty Ltd, Brisbane, Australia
[*]these authors contributed equally

Corresponding Author:
Elke Hacker, PhD
Institute of Health and Biomedical Innovation
School of Public Health and Social Work
Queensland University of Technology
60 Musk Avenue
Kelvin Grove
Brisbane, 4059
Australia
Phone: 61 731389674
Email: elke.hacker@qut.edu.au

Abstract

Background: Skin cancer is the most prevalent but also most preventable cancer in Australia. Outdoor workers are at increased risk of developing skin cancer, and improvements in sun protection are needed. Sunscreen, when applied at the recommended concentration (2 mg/cm^2), has been shown to block the harmful molecular effects of ultraviolet radiation in vivo. However, sunscreen is often not applied, reapplied sufficiently, or stored adequately to yield protection and reduce sunburns.

Objective: The primary aim of this study was to test an Internet of Things approach by deploying a smart sunscreen station to an outdoor regional mining site.

Methods: We deployed a smart sunscreen station and examined the key technological considerations including connectivity, security, and data management systems.

Results: The smart sunscreen station was deployed for 12 days at a mining workplace (Dalby, Australia). The smart sunscreen station's electrical components remained operational during field testing, and data were received by the message queuing telemetry transport server automatically at the end of each day of field testing (12/12 days, 100% connectivity).

Conclusions: This study highlights that an Internet of Things technology approach can successfully measure sunscreen usage and temperature storage conditions.

(J Med Internet Res 2020;22(5):e17079) doi:10.2196/17079

KEYWORDS

skin neoplasms; melanoma; health promotion; public health; preventive medicine; web applications

Introduction

Ultraviolet radiation (UVR) is the main environmental risk factor for melanoma and keratinocyte skin cancers. Australia has one of the highest rates of skin cancers in the world, which are at least double those of the United States or United Kingdom [1]. A meta-analysis has shown that outdoor workers are at increased risk of developing skin cancer, with the risk of squamous-cell carcinoma nearly double for outdoor workers than for indoor workers [2], while the risk of basal cell carcinoma is increased by almost 1.5 times [3]. In the Australian population, approximately 23.1% of all workers are estimated to be exposed to UVR at work [4]. Despite continuing educational efforts, a disconnect persists between public understanding of the harmful effects of excessive sun exposure and regular use of sunscreen. The Australian Work Exposures Study reported on 1100 outdoor workers and found that, although sun protection was used by 94.9% of Australian outdoor workers, only 8.7% of workers were classified as fully protected, with the most frequently used methods being

protective clothing and hats [4]. In an Australian outdoor workers study (n=162), 93% of workers reported sunscreen was provided by their workplace; however, sunscreen was the least frequently used personal protective equipment (PPE), with only 40% of workers self-reporting usually or always using sunscreen during work hours [5]. There was also a high percentage of outdoor workers (56.9%) exposed to UVR at work for more than 4 hours per day, with occupations such as mining, farming, and animal and horticultural workers [4]. Skin cancer has been listed as a work-related condition requiring priority for prevention activities in the Australian Work Health and Safety Strategy 2012-2022, and PPE is required to be provided by the workplace [6].

Regular sunscreen application has been shown to reduce the incidence of squamous cell carcinomas and melanoma [7] and block the harmful molecular effects of UVR on skin cells in vivo [8]. Health interventions to improve sun protection behaviors often assess sunscreen usage through self-report, which can be subject to social desirability bias and inaccurate memory recall [9,10]. Questionnaires or interviews can also be burdensome to complete, participants may miss questions, and they do not allow for real-time monitoring of compliance. Objective measurements of sunscreen usage are possible by using standardized scales and weighing bottles to determine the amount used. Limitations of the weighing method include the lack of a time stamp on usage data, reliance on the return of the sunscreen bottle for weighing, and inability to account for how many people may be using the sunscreen. Armstrong and colleagues [11] previously developed an electronic monitoring device to measure sunscreen usage, which involved a mobile phone being strapped to a sunscreen bottle. Each time the cap on the sunscreen tube was removed, the electronic monitor sent an SMS message in real time to a database where the dates and times of openings were recorded. However, the device was bulky to carry, and the database interface was limited [11].

The Internet of Things (IoT) is a concept where ordinary items are upgraded to include internet connectivity, allowing them to transmit information. In the health system, IoT-enabled smart devices have been used for preventative health in a US hospital to measure how often hospital workers were washing their hands [12]. The system had a ceiling-mounted sensor to monitor the flow of people as well as soap and hand sanitizer electronic monitoring dispensers to track usage. This provided hospital managers with a compliance rate of how many people washed their hands upon entering certain areas.

To improve sunscreen application in the workplace, this study aimed to leverage IoT-enabled smart devices by developing a smart sunscreen station that can stream data to an online management system to assist health and safety managers in determining the frequency of sunscreen use, measuring the temperature where the sunscreen is kept, and being alerted when sunscreen containers need replacing.

Methods

Smart Sunscreen Station Development

This study developed an IoT smart sunscreen station that objectively measures the use of a commercially available 1-L pump sunscreen product. The commercially available sunscreen product was retrofitted with magnets and flexible printed circuit board sensors to measure the pushing of the pump mechanism and dispensing of sunscreen (Figure 1). The sunscreen sensor assembly was then connected via a flat flex cable to a printed circuit board inside the smart sunscreen station housing that contained batteries, a main microcontroller (ESP32, Espressif, Shanghai, China), a cellular Cat-NB1 radio (BG96, Quectel Wireless Solutions Co Ltd, Shanghai, China), and processing electronics. The housing was constructed from plastic to allow the cellular antennas to be mounted internally to protect them from exposure to the environment (Figure 2). Cellular connectivity was provided via a Cat-NB1 network (Telstra Corporation Limited, Melbourne, Australia). An open-source message queuing telemetry transport broker/server (Mosquitto, Eclipse Foundation, Ottawa, Canada) was used to collect the data from the device over the internet (Figure 3), and all communication between the device and server was encrypted using Transaction Layer Security (TLS) version 1.2 (Figure 4). Data were stored in a SQL database and served using a small web application written in Python v3.

Figure 1. Smart sunscreen station measuring sunscreen dispensing with (A) retrofitted sunscreen sensor assembly and (B) pump mechanism and sensor detection concept. PCB: printed circuit board.

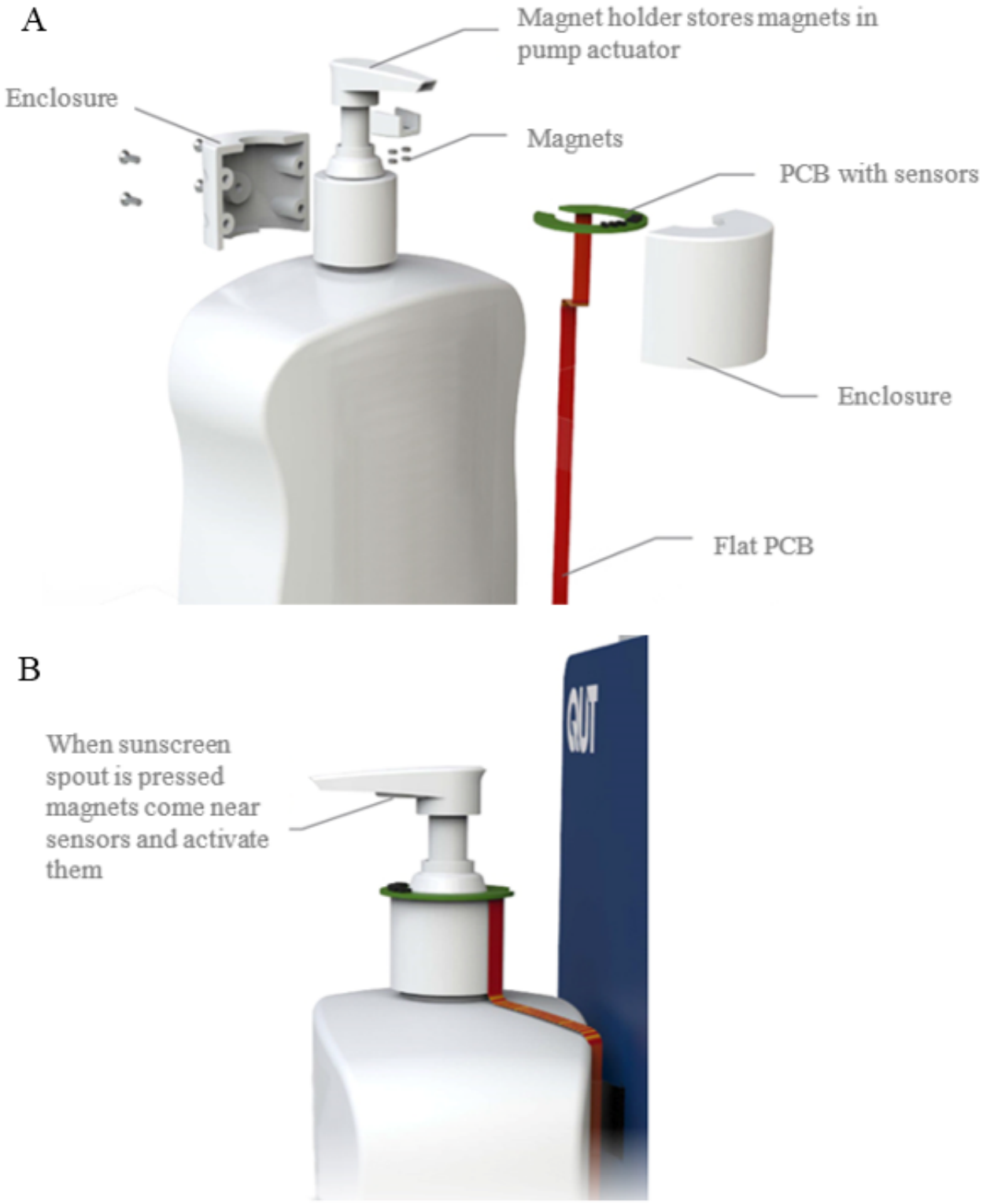


Figure 2. Smart sunscreen station housing with electrical components housed within the bottom of the stand and stainless steel strapping to secure the bottle in place. PCB: printed circuit board.

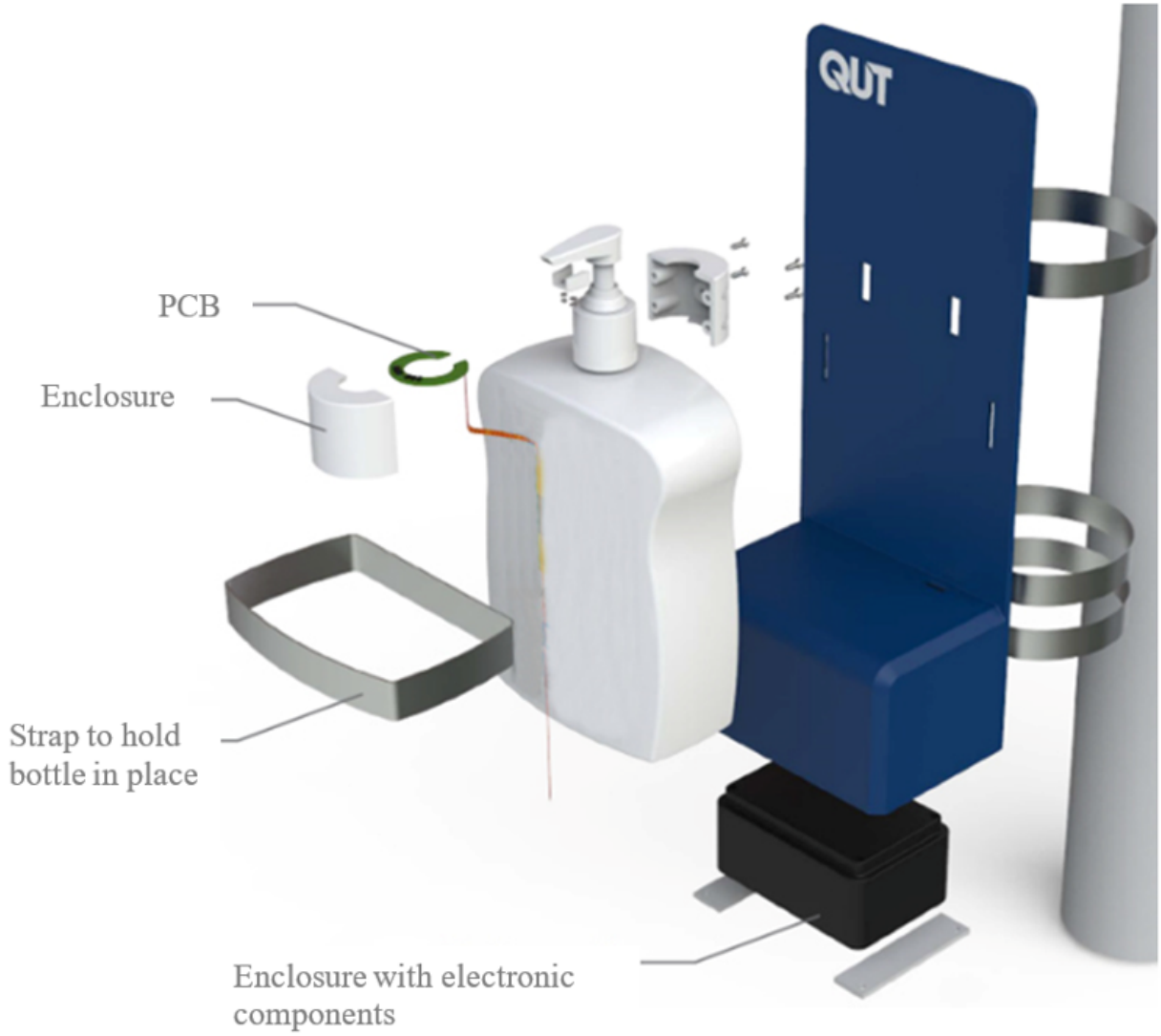


Figure 3. Smart sunscreen station electronic components; the battery housing was protected from water and dust using rubber seals (A), and the BG96 module was connected to a narrowband Internet of Things subscriber identification module (SIM) card and located underneath the battery platform with one NB1 radio antenna attached to the wall of the housing (B).

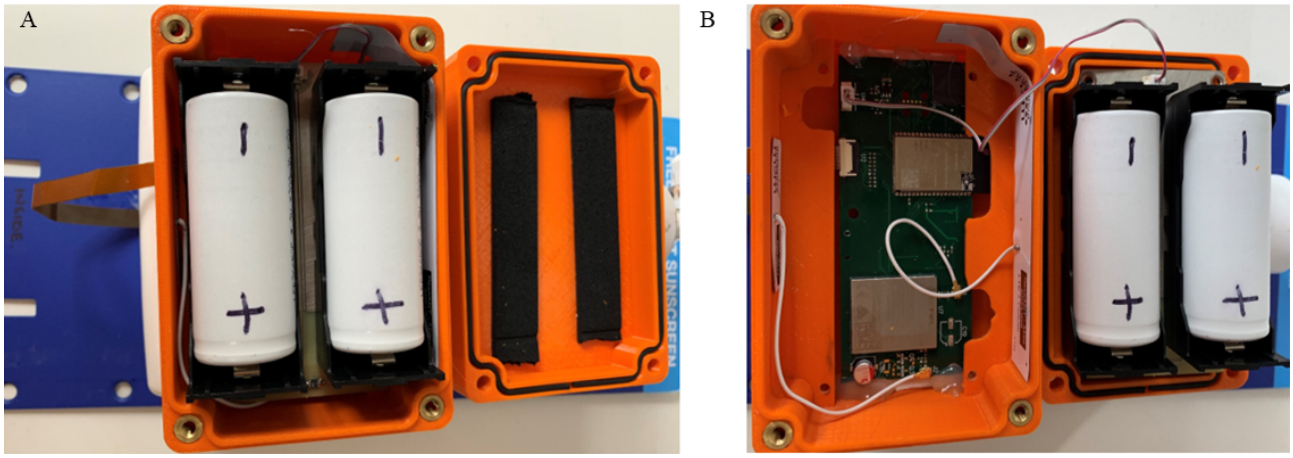
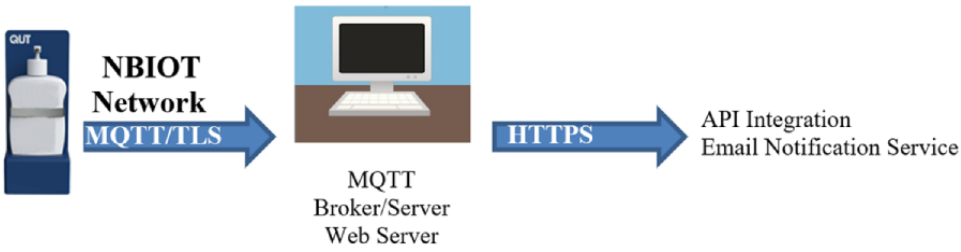


Figure 4. The process flow of data from an Internet of Things (IoT) smart sunscreen station. API: application programming interface; MQTT: message queuing telemetry transport; NBIOT: narrowband Internet of Things; TLS: Transaction Layer Security.



Observational Laboratory Testing

To check the smart sunscreen station was connecting and recording data to the server correctly, observational testing was performed in Brisbane, Australia (approximate latitude 27 ^{o}S, 153 ^{o}E). A full 1-L sunscreen bottle was pumped 10 times, and the bottle was weighed to confirm the volume of sunscreen dispensed. A half-empty 1-L sunscreen bottle with a net weight of 500 g was pumped 10 times and weighed. A nearly empty 1-L sunscreen bottle with a net weight of 200 g was pumped 10 times and weighed. The time-stamped data on the server were then compared with the observed time-recorded pumps and weight of the bottle.

The smart sunscreen station underwent heat testing by placing the unit in an oven at 35 ^{o}C for 1 hour. The temperature of the smart sunscreen station was recorded using an infrared handheld thermometer (ThermaTwin TN410LCE Infrared Thermometer, OneTemp, Adelaide, Australia). The smart sunscreen station was further heat tested by placing it in an oven at 45 ^{o}C for 30 minutes, and data were recorded for an additional 12 days.

Smart Sunscreen Station Deployment

The field study was conducted during November 2018 to December 2018 (summer in Australia) when the UV Index in Queensland is consistently above 9 and can reach 14+, requiring sun protection every day. One smart sunscreen station was deployed to an outdoor workplace located in Dalby, Australia (latitude 27 ^{o}S, 151 ^{o}E). The smart sunscreen station was taken out to drilling rig sites located between Dalby and Chinchilla during the testing period of November 25, 2018 to December 6, 2018. The site was an operational gas mining venture that had worker activity in a high UV environment and no underground workforce. The smart sunscreen station was placed in the air-conditioned meeting and lunch room, which is a high traffic area that is accessed by all workers at the start of the day and during breaks. This is where existing safety talks are held before commencing outdoor work. The sunscreen used in this study was commercially available SPF 50+ (the highest SPF level available in Australia). The smart sunscreen station did not monitor personal use of sunscreen or who had used it. It measured the number of times the sunscreen was pumped and the storage temperature, and it sent notifications to the database when it required sunscreen replacement.

Feedback was obtained from end users including health and safety officers, purchasing officers, and health officers to design and refine the software dashboard. Questions surrounding the dashboard infographics were asked to assist with communicating the data collected on the IoT smart sunscreen station. In addition, email and phone contact details of the researchers were provided to the workers for complaints, technical issues, or further information during the deployment. The deployment of the smart sunscreen station was to assess the functionality and not human subjects' research; therefore, we obtained institutional ethics review board exemption.

Statistical Analysis

Sunscreen usage variables were dichotomized to categorical data: Sunscreen usage data were coded "yes" if the timestamped data of sunscreen use were recorded to the server and coded "no" if the sunscreen bottle was pumped but data were not recorded to the server. The Cohen kappa score was calculated to determine if there was agreement between categorical variables for sunscreen usage. Sunscreen weight data were recorded in grams, and the observed weight per pump was calculated and compared to the server-predicted weight. Spearman's rank correlation coefficient was used to determine correlations between server-predicted and observed weight. Values >0.4-0.6 were considered moderate, >0.6-0.8 substantial, and >0.8-1.0 almost perfect agreement. Prism graph-pad (version 7, GraphPad Software, San Diego, CA) and SPSS software (version 25.0, IBM Corp, Armonk, NY) were used for analyses.

Results

Laboratory Testing and Connectivity

Laboratory testing of the smart sunscreen station demonstrated there was perfect agreement between observed sunscreen use and server-recorded sunscreen use (κ=1.0, 95% CI 1.00-1.00; Table 1). There was almost perfect agreement between the observed weight of the sunscreen bottle and predicted weight based off server usage data (r=1.0, 95% CI 1.00-1.00, *P*<.001; Table 1). The smart sunscreen station temperature sensor recorded to the server that the temperature exceeded the maximum storage temperature of 30 ^{o}C following a 1-hour oven incubation at 35 ^{o}C. The unit was tested further with a 1-hour oven incubation at 45 ^{o}C, and the unit continued to be operational and connect to the server and send data during the post-heat stress measurement period (100% connectivity, 12/12 days).

Table 1. Agreement between observed sunscreen usage and server-recorded measurements.

Sunscreen bottle status	Sunscreen pump recorded in the server		Observed average weight per pump (g)	Server-predicted weight per pump (g)
	Yes, n (%)	No, n (%)		
Full bottle weighing 1000 g bottle (n=10 pumps)	10 (100)	0 (0%)	1.732	1.60
Half-empty bottle weighing 500 g (n=10 pumps)	10 (100)	0 (0%)	1.564	1.55
Nearly empty bottle weighing 200 g (n=10 pumps)	10 (100)	0 (0%)	1.526	1.50

Field Testing and Connectivity

The smart sunscreen station uploaded data each day to the message queuing telemetry transport server during the outdoor workplace deployment (Figure 5). No complaints nor concerns were logged from the workforce during the 12 days the smart sunscreen station was deployed. The smart sunscreen station was kept in an air-conditioned site meeting and lunch room at a drill rig site and did not exceed the maximum storage temperature of 30 ^{o}C during the field test. The network coverage for the region where the smart sunscreen station was located illustrated good reception as shown by the green areas in the map (Figure 6). The smart sunscreen station recorded data locally each day, and data were uploaded to the server once each evening as programmed by the software, with data received for 12 of the 12 days (100% connectivity). The smart sunscreen station did not require continuous connectivity and stored data locally, only transmitting data once a day. The data transmission took approximately 10 seconds, with a 70-mA average transmit current as measured on the device. Sleep current was measured at approximately 10-20 uA. As a result, in normal conditions, the system battery life is expected to be predominantly limited by self-discharge of the lithium iron phosphate ($LiFePO_4$) batteries.

Figure 5. Data received each day from the smart sunscreen station, including number of pumps and temperature status.

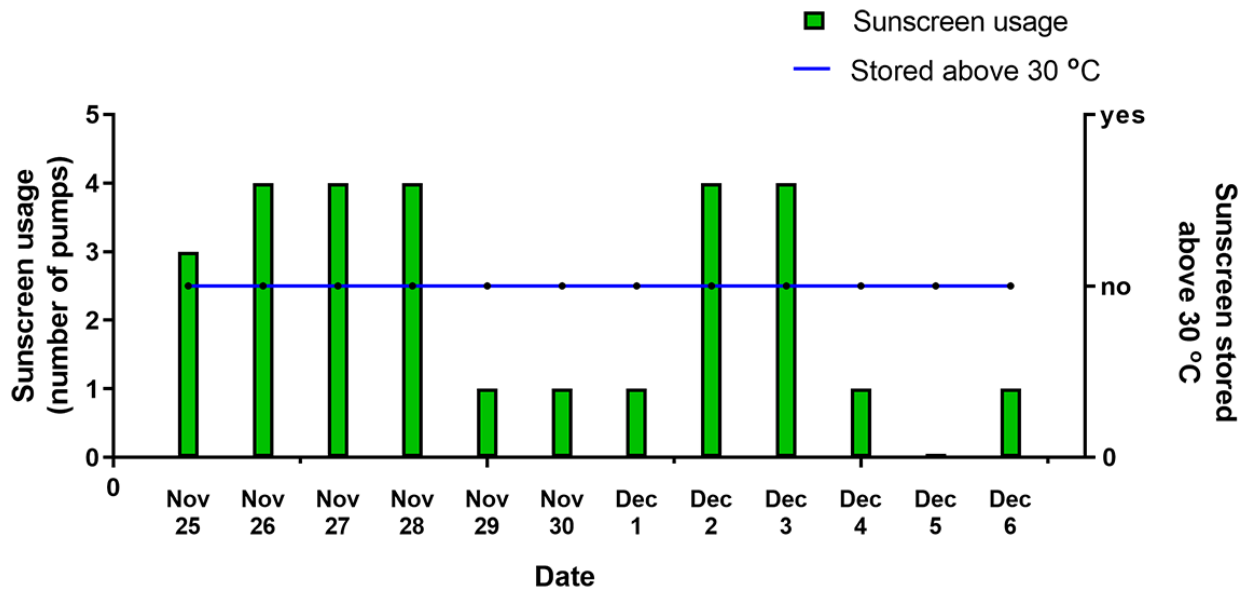
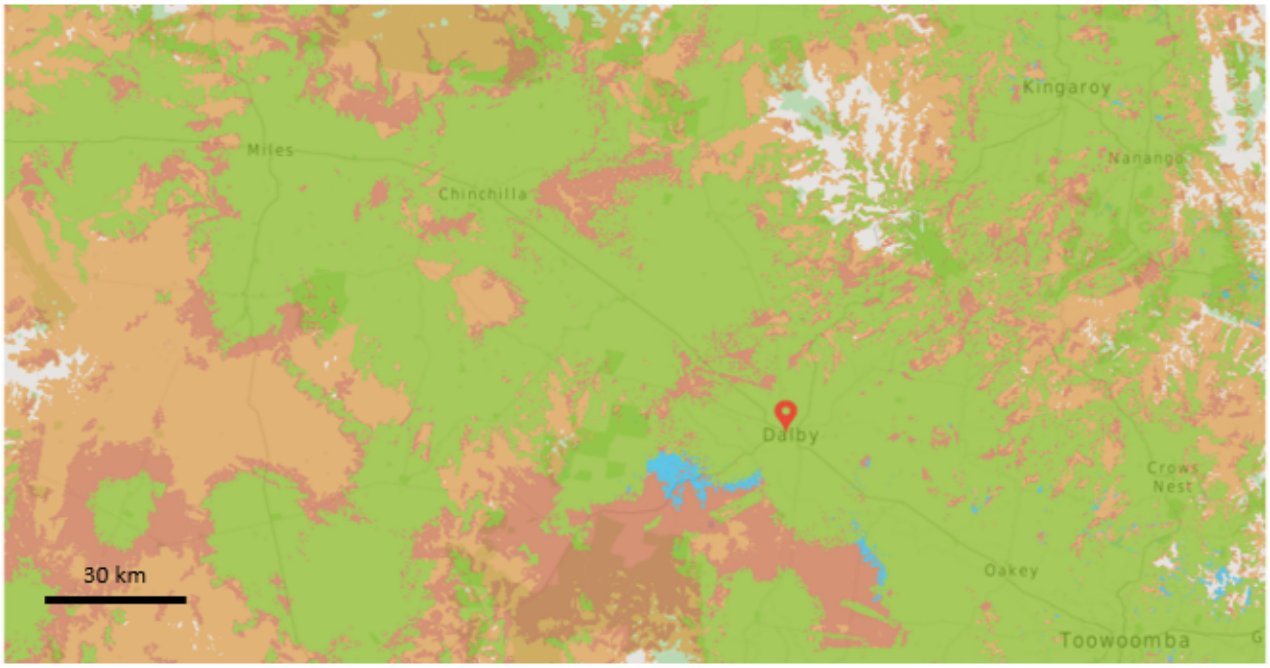


Figure 6. Network coverage for the field testing site in Dalby, Queensland (green, good 4G reception; blue, slower 4G reception; red, 3G reception only; orange, external antenna required for reception). Image adapted from [13].



Security

All communications over the internet were protected using TLS v1.2 with a preshared certificate in the firmware of the device. Encryption was performed on the microcontroller itself to prevent the data being read or modified using a man-in-the-middle attack between the microcontroller and the cellular radio. In addition, the microcontroller utilized flash encryption and secure boot to prevent any unauthorized readout or modification of the firmware.

Data Management Systems

In consultation with stakeholders from health and safety roles at outdoor workplaces (n=5), dashboard infographics were developed to assist with communicating the data collected on the smart sunscreen station (Figures 7-9). Sunscreen usage data were converted into consumer-friendly infographics with the number of sunscreen pumps plotted over time using the time-stamp information collected. Historical data for previous months can be displayed and overlaid with color-coded UV Index weather information to illustrate trends in sunscreen usage (Figure 7). The level of sunscreen remaining in each bottle can be estimated using the number of pumps multiplied by the weight dispensed, which is calculated by the data management system and converted into an infographic (Figure 8). The level of sunscreen can then be plotted over time to assist with maintenance scheduling, and notifications are sent to alert users when sunscreen bottles need to be changed (Figures 8 and 9).

Figure 7. Dashboard infographics communicating smart sunscreen station data, including the daily sunscreen usage per person and the monthly sunscreen usage overlaid with UV index weather data.

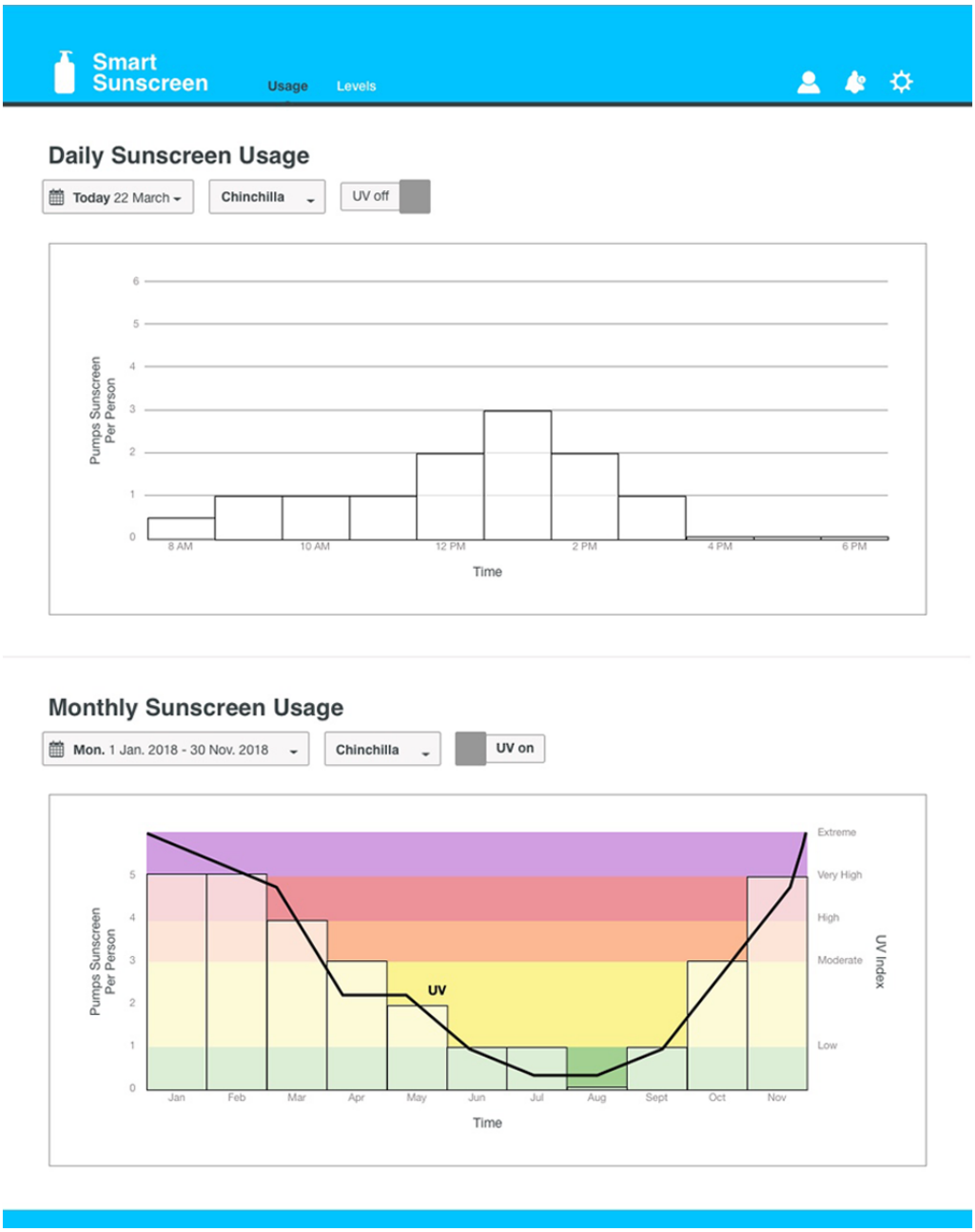


Figure 8. Smart sunscreen station dashboard data showing the current sunscreen bottle level the level of sunscreen in the bottle over time.

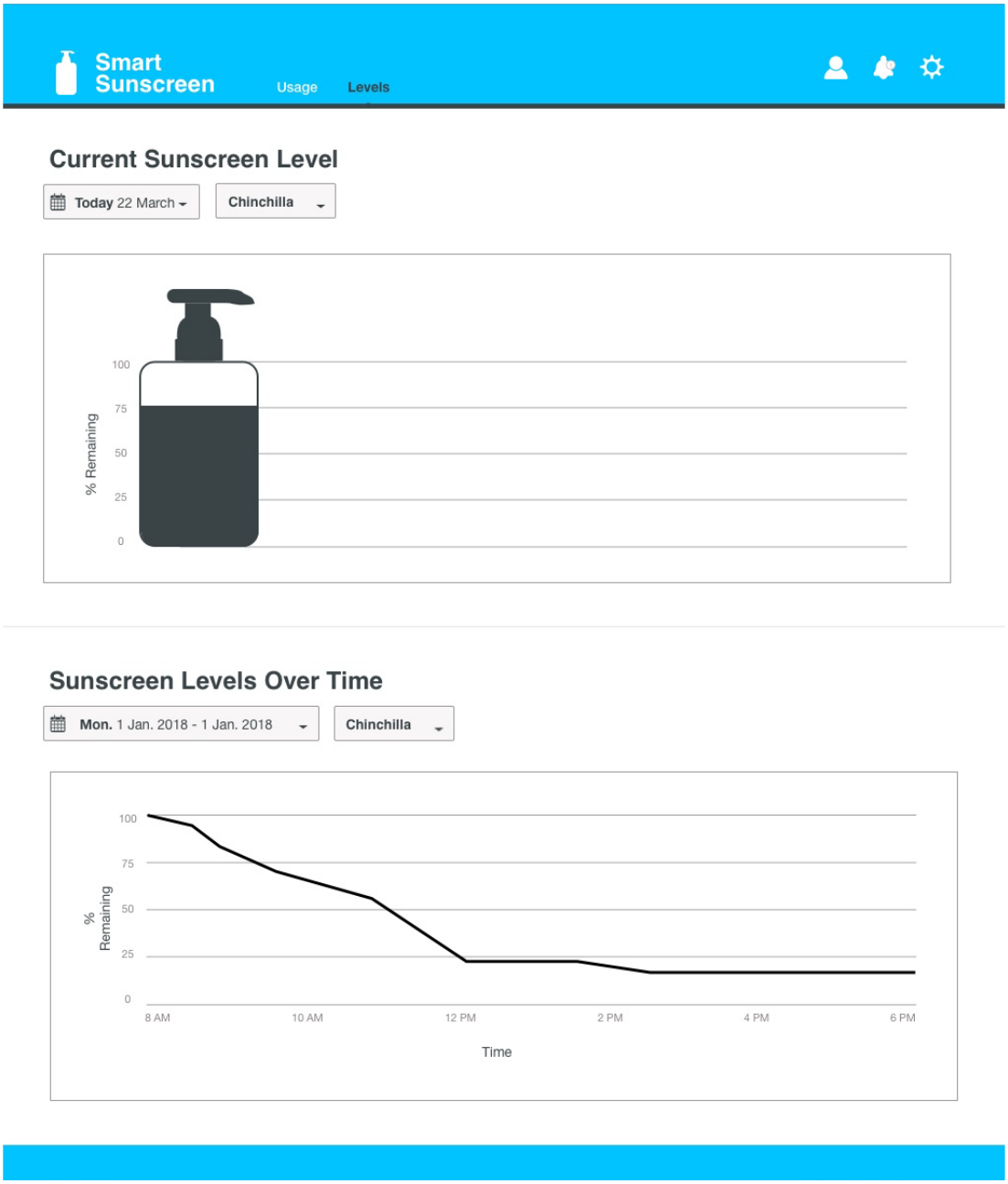
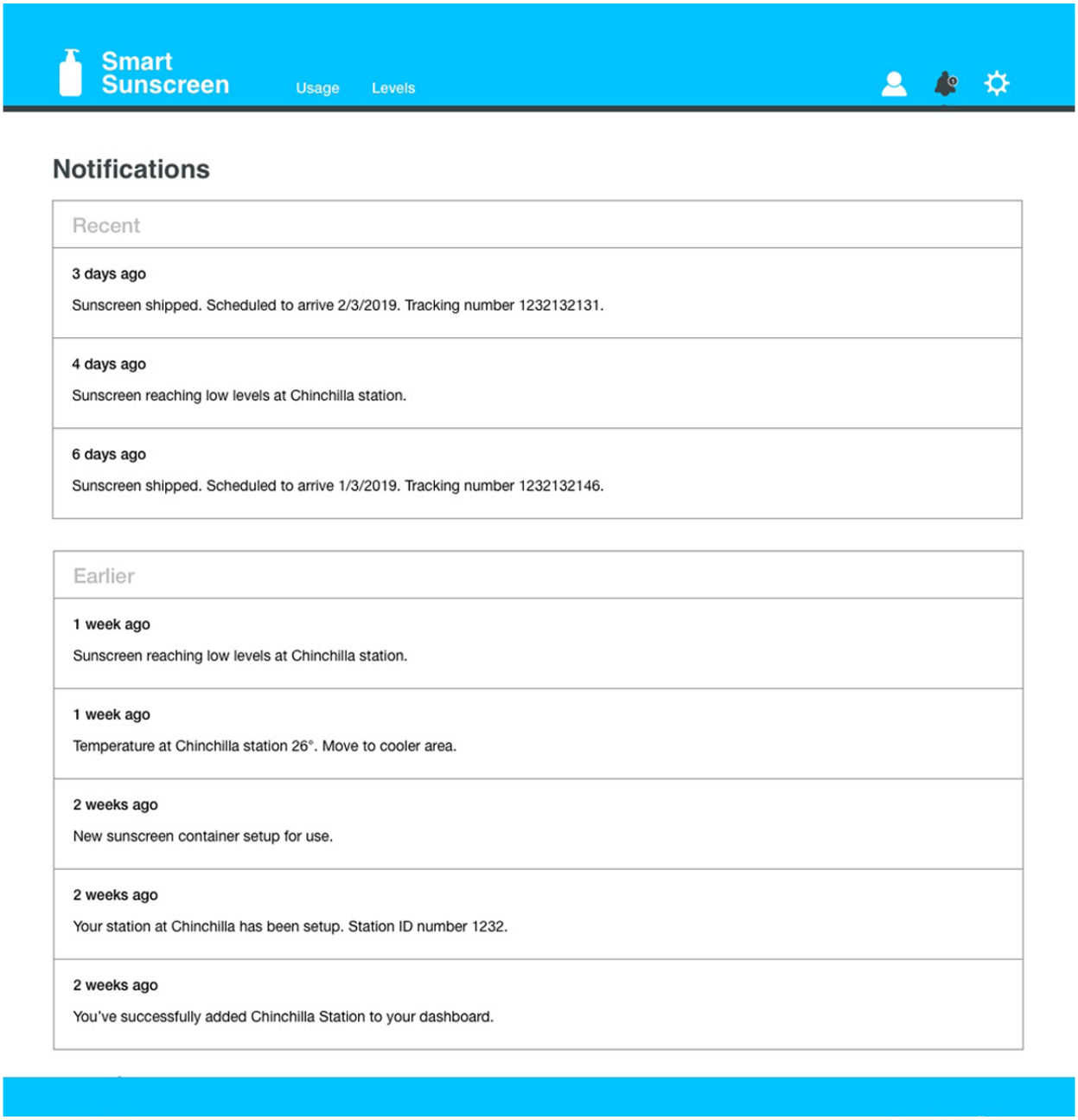


Figure 9. Smart sunscreen station dashboard showing notifications and alerts.



Discussion

Principal Findings

Our data demonstrated that this IoT smart sunscreen station can accurately record time-stamped sunscreen use from a 1-L bottle in an outdoor worker setting. During the 12-day field testing deployment in an outdoor workplace in regional Australia, the station's electrical components remained operational, and data were received at the server each day, resulting in 100% connectivity. In addition, we reported a high correlation between observed sunscreen use and data recorded in the server as well as the observed weight of the sunscreen bottle and the predicted weight generated by the software based from the server usage data. Based on the study settings and results, we suggest that the smart sunscreen station may become a valuable tool to help optimize sunscreen supply and maintenance in an outdoor workplace.

Outdoor workers are at risk of developing skin cancer and are typically not using adequate sun protection, with studies showing that workers who perceived sun protection as valued in their workplace had higher levels of sun protection [14,15]. Uptake of sun protection by outdoor workers is affected by both workplace and personal factors, and strategies should target the workplace environment and workers' attitudes and behaviors [5].The provision of PPE by the workplace has been suggested to improve sun protection in outdoor workers by providing broader support to promote positive behaviors [16]. The smart sunscreen station shows significant promise to generate relevant data for workplace health and safety legislative requirements by illustrating that appropriate PPE is being supplied to workers. The real-time monitoring and inventory management of sunscreen would ensure appropriate levels of sunscreen are

always available to workers, allow targeted training as part of health and safety programs in low usage areas, and ensure the sunscreen provided has not been overheated leading to reduced effectiveness.

Previous research has found low usage of sunscreen stations in public settings [17]. Kirby et al [17] implemented a free sunscreen station (which did not have IoT technology) at an amusement park, and the researchers observed usage from a distance by counting the number of people who passed, approached, or interacted with the sunscreen station. During the 3 observation periods, 879 people passed the sunscreen station, 2.6%-5.1% of visitors showed interest in the station, and even fewer visitors (0%-2.9%) used the sunscreen station. This study by Kirby et al [17] reported that members of the public had trouble recognizing the sunscreen station and suggested increased signage. In the outdoor workplace setting, sunscreen forms an important PPE measure, and data on sunscreen usage could be used to inform workforce PPE compliance and practices. A British study of outdoor workers reported that sun protection education was associated with workers' use of sun protective measures [18]. Other barriers to sunscreen use include that it is time-consuming and wearing it may be uncomfortable due to greasiness or increased sweatiness. Testing a variety of sunscreens using the smart sunscreen station to determine a preferred type could increase sunscreen use.

The IoT technology would also enable objective data collection rather than relying on self-report or conducting observational research in settings evaluating health promotion programs targeted to increase sun protection habits. Future studies will enable not only outdoor workplaces but also other high UV environments such as recreational venues and schools to evaluate sunscreen usage.

Ensuring secure and reliable network connectivity is an important consideration in IoT deployments, particularly in rural or regional areas. Unlike streaming applications, the IoT smart sunscreen station does not require continuous connectivity, as the device only needs to transmit data once per day; therefore, battery life can be conserved, and the servicing requirements of the IoT device are minimal. However, the data transmission is versatile and can be adjusted to range from seconds to days. For example, in public spaces or workplaces with a high volume of passing traffic, 5-minute data transmission intervals could be programmed, and this would allow replacement of sunscreen in a timely manner.

The narrowband IoT network was able to provide wide coverage when compared to other network alternatives. During this study, we demonstrated reliable connectivity in diverse areas from city centers to regional districts. The advancements in publicly available cloud-based services such as Amazon Web Services and Microsoft Azure for IoT devices also provide a cost-competitive alternative to traditional, on-premise hosting environments. These platforms allow users to connect their assets using industry standard security systems (such as TLS) to safely and securely gather data and undertake bidirectional communication.

The backend data management system processed the data received from the IoT device and was an important part of the design and crucial in allowing the data to be interpreted by the end user. This study found the use of infographics to convey IoT data to end users was acceptable and preferred over tables. These results are similar to other reports in health care communication, which show that engaging methods of visually communicating information in a colorful and concise manner are preferred [19]. The smart sunscreen device may utilize a lease arrangement with the options of 12-month or 24-month subscriptions, which include servicing and maintenance of the units and monthly scheduling of fees. This reduces the upfront prohibitive cost for the workplace to purchase the smart sunscreen station outright.

IoT has the potential to collect and integrate data in real time to improve operational efficiencies. However, to our knowledge, no previous studies have investigated the effective integration and application of IoT for sunscreen monitoring in an outdoor workplace. Previous work by Wu and colleagues [20] using an IoT wearable body sensor measured body temperature, heart rate, and environmental conditions such as temperature, carbon dioxide, UV Index, and humidity, and the data collected were used to trigger an alert if any emergency situation occurred. The body sensor data were connected using a low-power wide-area network, and data transmission was restricted to occur only within the gateway network area, which was reported to be 520 m indoors and 926 m when outdoors. The study was based on construction workers, reported high UVR exposure, and demonstrated the potential to gather safety monitoring information to assist workplace health and safety programs [20]. IoT technology has also previously been deployed as a notification system to measure restroom cleanliness in hospitals, whereby public users press the button when the restroom requires cleaning [21]. Chai and colleagues [21] reported this was a feasible method to streamline maintenance activities like restroom cleaning and would be further extended to include critical supply restocking tasks. Unlocking the potential for IoT technologies to assist and lead to evidence-based improvements requires evaluation of the impact of devices as well as strategies to engage end users and communicate the benefits yielded from the data collected. Key technical considerations for a successful deployment include ensuring appropriate network connectivity, designing a product with a robust management system that operates with current organizational platforms, planning for storage in high UV environments, and carefully considering the configuration to minimize privacy and security risks.

The aim of this study was to assess the proof of concept of a smart sunscreen station in a regional outdoor workplace. Limitations of the study include that we did not analyze the software data nor capture data on health behavior change such as the sun protection habit index or employee engagement; in addition, the duration of field testing was only 12 days. The interview analysis was limited by a small sample size (n=5), and future research could expand on additional enablers and barriers to sunscreen use by outdoor workers. Future projects deploying IoT devices could also explore further implementation factors including human aspects such as attitudes, behavior, and capacity as well as organizational, financial, and legislative processes.

Conclusion

Consistent use of sunscreen is recommended to prevent sunburn and reduce the risk of developing skin cancers. Sunscreen is an essential PPE for outdoor workers, particularly on the face, neck, and hands. To assist workplaces to protect their workers from UVR, the smart sunscreen station could help streamline supply, restocking, and storage requirements, especially in high-temperature environments, as well as benefit health and safety programs with targeted training. This study provides evidence for the technical feasibility of IoT smart sunscreen stations in outdoor workplaces including regional and remote areas of Australia.

Acknowledgments

The authors would like to thank the outdoor workplace QGC-Shell Australia for their support to deploy the device and Telstra Corporation Limited for access to pilot test SIM cards via their Cat-NB1 network. The sponsors of the study (Advance Queensland) had no role in the study design, collection, analysis, and interpretation of data; in the writing of this manuscript; and in the decision to submit the paper for publication. The corresponding author had full access to all data in the study and final responsibility for the decision to submit for publication.

Conflicts of Interest

Authors HF, CH, and EH state no conflict of interest. Authors JH and AW are employees of Designworks group.

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 2015 Mar 01;136(5):E359-E386 [FREE Full text] [doi: 10.1002/ijc.29210] [Medline: 25220842]
2. Schmitt J, Seidler A, Diepgen TL, Bauer A. Occupational ultraviolet light exposure increases the risk for the development of cutaneous squamous cell carcinoma: a systematic review and meta-analysis. Br J Dermatol 2011 Feb;164(2):291-307. [doi: 10.1111/j.1365-2133.2010.10118.x] [Medline: 21054335]
3. Bauer A, Diepgen TL, Schmitt J. Is occupational solar ultraviolet irradiation a relevant risk factor for basal cell carcinoma? A systematic review and meta-analysis of the epidemiological literature. Br J Dermatol 2011 Sep;165(3):612-625. [doi: 10.1111/j.1365-2133.2011.10425.x] [Medline: 21605109]
4. Carey RN, Glass DC, Peters S, Reid A, Benke G, Driscoll TR, et al. Occupational exposure to solar radiation in Australia: who is exposed and what protection do they use? Aust N Z J Public Health 2014 Feb;38(1):54-59. [doi: 10.1111/1753-6405.12174] [Medline: 24494947]
5. Janda M, Stoneham M, Youl P, Crane P, Sendall MC, Tenkate T, et al. What encourages sun protection among outdoor workers from four industries? J Occup Health 2014;56(1):62-72 [FREE Full text] [doi: 10.1539/joh.13-0179-oa] [Medline: 24270927]
6. Safe Work Australia. 2018. Australian Work Health and Safety Strategy 2012-2022 URL: https://www.safeworkaustralia.gov.au/about-us/australian-work-health-and-safety-strategy-2012-2022 [accessed 2019-06-05]
7. Green AC, Williams GM, Logan V, Strutton GM. Reduced melanoma after regular sunscreen use: randomized trial follow-up. J Clin Oncol 2011 Jan 20;29(3):257-263. [doi: 10.1200/JCO.2010.28.7078] [Medline: 21135266]
8. Hacker E, Boyce Z, Kimlin MG, Wockner L, Pollak T, Vaartjes SA, et al. The effect of MC1R variants and sunscreen on the response of human melanocytes in vivo to ultraviolet radiation and implications for melanoma. Pigment Cell Melanoma Res 2013 Nov;26(6):835-844. [doi: 10.1111/pcmr.12157] [Medline: 23962207]
9. Hacker E, Horsham C, Vagenas D, Jones L, Lowe J, Janda M. A Mobile Technology Intervention With Ultraviolet Radiation Dosimeters and Smartphone Apps for Skin Cancer Prevention in Young Adults: Randomized Controlled Trial. JMIR Mhealth Uhealth 2018 Nov 28;6(11):e199 [FREE Full text] [doi: 10.2196/mhealth.9854] [Medline: 30487115]
10. Glanz K, McCarty F, Nehl EJ, O'Riordan DL, Gies P, Bundy L, et al. Validity of self-reported sunscreen use by parents, children, and lifeguards. Am J Prev Med 2009 Jan;36(1):63-69 [FREE Full text] [doi: 10.1016/j.amepre.2008.09.012] [Medline: 18945582]
11. Armstrong AW, Watson AJ, Makredes M, Frangos JE, Kimball AB, Kvedar JC. Text-message reminders to improve sunscreen use: a randomized, controlled trial using electronic monitoring. Arch Dermatol 2009 Nov;145(11):1230-1236. [doi: 10.1001/archdermatol.2009.269] [Medline: 19917951]
12. Edmonds S, Brown M, White C, Kirk JM. Impact of Electronic Monitoring and a Hand Hygiene Improvement Program on Compliance Rates. Am J Infection Control 2013;41(6):S38.
13. Telstra Corporation Limited, 2019. Check the coverage available in your area URL: https://www.telstra.com.au/coverage-networks/our-coverage [accessed 2019-09-13]
14. McCool JP, Reeder AI, Robinson EM, Petrie KJ, Gorman DF. Outdoor workers' perceptions of the risks of excess sun-exposure. J Occup Health 2009;51(5):404-411 [FREE Full text] [doi: 10.1539/joh.l9030] [Medline: 19590155]

15. Hammond V, Reeder AI, Gray AR, Bell ML. Are workers or their workplaces the key to occupational sun protection? Health Promot J Austr 2008 Aug;19(2):97-101. [doi: 10.1071/he08097] [Medline: 18647121]
16. Horsham C, Auster J, Sendall MC, Stoneham M, Youl P, Crane P, et al. Interventions to decrease skin cancer risk in outdoor workers: update to a 2007 systematic review. BMC Res Notes 2014 Jan 07;7:10 [FREE Full text] [doi: 10.1186/1756-0500-7-10] [Medline: 24397996]
17. Kirby JS, Hansen TJ, Sabella C, Sisti AJ, Fields M. Summertime scorcher: assessing and promoting sunscreen protection in an amusement park setting. Photodermatol Photoimmunol Photomed 2014 Aug;30(4):195-201. [doi: 10.1111/phpp.12110] [Medline: 24417478]
18. Madgwick P, Houdmont J, Randall R. Sun safety measures among construction workers in Britain. Occup Med (Lond) 2011 Sep;61(6):430-433. [doi: 10.1093/occmed/kqr054] [Medline: 21752939]
19. McCrorie AD, Donnelly C, McGlade KJ. Infographics: Healthcare Communication for the Digital Age. Ulster Med J 2016 May;85(2):71-75 [FREE Full text] [Medline: 27601757]
20. Wu F, Wu T, Yuce MR. An Internet-of-Things (IoT) Network System for Connected Safety and Health Monitoring Applications. Sensors (Basel) 2018 Dec 21;19(1):e21 [FREE Full text] [doi: 10.3390/s19010021] [Medline: 30577646]
21. Chai PR, Zhang H, Jambaulikar GD, Boyer EW, Shrestha L, Kitmitto L, et al. An Internet of Things Buttons to Measure and Respond to Restroom Cleanliness in a Hospital Setting: Descriptive Study. J Med Internet Res 2019 Jun 19;21(6):e13588 [FREE Full text] [doi: 10.2196/13588] [Medline: 31219046]

Abbreviations

API: application programming interface.
IoT: Internet of Things.
MQTT: message queuing telemetry transport.
NBIoT: narrowband Internet of Things.
PCB: printed circuit board.
PPE: personal protective equipment.
TLS: Transaction Layer Security.
UVR: ultraviolet radiation.

Edited by G Eysenbach; submitted 22.11.19; peer-reviewed by G Jambaulikar, M Sendall, P Wickner; comments to author 12.12.19; revised version received 21.12.19; accepted 29.03.20; published 28.05.20.

Please cite as:
Ford H, Herbert J, Horsham C, Wall A, Hacker E
Internet of Things Smart Sunscreen Station: Descriptive Proof-of-Concept Study
J Med Internet Res 2020;22(5):e17079
URL: http://www.jmir.org/2020/5/e17079/
doi:10.2196/17079
PMID:32463378

Corrigenda and Addenda

Correction: The Effect of Narrative on Physical Activity via Immersion During Active Video Game Play in Children: Mediation Analysis

Caio Victor Sousa[1], PhD; Austin Fernandez[1], BSc; Jungyun Hwang[2], PhD; Amy Shirong Lu[1], PhD

[1]College of Arts, Media, and Design, Bouvé College of Health Sciences, Health Technology Lab, Northeastern University, Boston, MA, United States
[2]Stanford University Medical Center, Department of Medicine, Palo Alto, CA, United States

Corresponding Author:
Amy Shirong Lu, PhD
College of Arts, Media, and Design, Bouvé College of Health Sciences
Health Technology Lab
Northeastern University
360 Huntington Ave
212A Lake Hall
Boston, MA, 02115
United States
Phone: 1 6173735795
Email: a.lu@northeastern.edu

Related Article:

Correction of: https://www.jmir.org/2020/3/e17994/

(J Med Internet Res 2020;22(5):e20134) doi:10.2196/20134

In the original published paper "The Effect of Narrative on Physical Activity via Immersion During Active Video Game Play in Children: Mediation Analysis" (J Med Internet Res 2020;22(3):e17994), the authors noticed an error in the Abstract and Table 2. This was caused by an updated analysis not being accurately reflected in the manuscript. The Abstract and Table 2 listed incorrect values of moderate-to-vigorous physical activity (MVPA) for the narrative (NV) group.

The value in the original Abstract section was initially listed as:

> *The NV group had significantly higher narrative immersion (mean 3.50, SD 0.55 vs mean 2.91, SD 0.59; P=.03) and MVPA than the NNV group (mean 20.11, SD 13.75 vs mean 7.85, SD 5.83; P=.02).*

The correct value is:

> *The NV group had significantly higher narrative immersion (mean 3.50, SD 0.55 vs mean 2.91, SD 0.59; P=.03) and MVPA (mean 19.46, SD 13.31 vs mean 7.85, SD 5.83; P=.02) than the NNV group.*

Furthermore, in Table 2, the "Narrative (n=12), mean (SD)" value for "Moderate-to-vigorous physical activity (minutes)" was originally listed as:

> *19.4 (13.3)*

The correct "Narrative (n=12), mean (SD)" value for "Moderate-to-vigorous physical activity (minutes)" in Table 2 is:

> *19.5 (13.3)*

The changes were not significant, and do not affect the overall findings of the paper.

The authors also want to acknowledge FableVision Studios contribution. The following sentence has been added to the Acknowledgments:

> *FableVision Studios produced the narrative video for this project.*

The correction will appear in the online version of the paper on the JMIR website on May 28, 2020, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 11.05.20; this is a non–peer-reviewed article;accepted 13.05.20; published 28.05.20.

Please cite as:
Sousa CV, Fernandez A, Hwang J, Lu AS
Correction: The Effect of Narrative on Physical Activity via Immersion During Active Video Game Play in Children: Mediation Analysis
J Med Internet Res 2020;22(5):e20134
URL: http://www.jmir.org/2020/5/e20134/
doi:10.2196/20134
PMID:32463788

Original Paper

Chinese Public's Attention to the COVID-19 Epidemic on Social Media: Observational Descriptive Study

Yuxin Zhao[1], MD; Sixiang Cheng[1], PhD; Xiaoyan Yu[1], MD; Huilan Xu[1], PhD

Department of Social Medicine and Health Management, Xiangya School of Public Health, Central South University, Changsha, China

Corresponding Author:
Huilan Xu, PhD
Department of Social Medicine and Health Management
Xiangya School of Public Health
Central South University
Xiangya School of Public Health, Central South University
Changsha, 410078
China
Phone: 86 13637480958
Email: xhl_csu@163.com

Abstract

Background: Since the coronavirus disease (COVID-19) epidemic in China in December 2019, information and discussions about COVID-19 have spread rapidly on the internet and have quickly become the focus of worldwide attention, especially on social media.

Objective: This study aims to investigate and analyze the public's attention to events related to COVID-19 in China at the beginning of the COVID-19 epidemic (December 31, 2019, to February 20, 2020) through the Sina Microblog hot search list.

Methods: We collected topics related to the COVID-19 epidemic on the Sina Microblog hot search list from December 31, 2019, to February 20, 2020, and described the trend of public attention on COVID-19 epidemic-related topics. ROST Content Mining System version 6.0 was used to analyze the collected text for word segmentation, word frequency, and sentiment analysis. We further described the hot topic keywords and sentiment trends of public attention. We used VOSviewer to implement a visual cluster analysis of hot keywords and build a social network of public opinion content.

Results: The study has four main findings. First, we analyzed the changing trend of the public's attention to the COVID-19 epidemic, which can be divided into three stages. Second, the hot topic keywords of public attention at each stage were slightly different. Third, the emotional tendency of the public toward the COVID-19 epidemic-related hot topics changed from negative to neutral, with negative emotions weakening and positive emotions increasing as a whole. Fourth, we divided the COVID-19 topics with the most public concern into five categories: the situation of the new cases of COVID-19 and its impact, frontline reporting of the epidemic and the measures of prevention and control, expert interpretation and discussion on the source of infection, medical services on the frontline of the epidemic, and focus on the worldwide epidemic and the search for suspected cases.

Conclusions: Our study found that social media (eg, Sina Microblog) can be used to measure public attention toward public health emergencies. During the epidemic of the novel coronavirus, a large amount of information about the COVID-19 epidemic was disseminated on Sina Microblog and received widespread public attention. We have learned about the hotspots of public concern regarding the COVID-19 epidemic. These findings can help the government and health departments better communicate with the public on health and translate public health needs into practice to create targeted measures to prevent and control the spread of COVID-19.

(J Med Internet Res 2020;22(5):e18825) doi:10.2196/18825

KEYWORDS

COVID-19; Sina Microblog; public attention; social media; China; public health emergency; infodemic

Introduction

The coronavirus disease (COVID-19) is a new infectious disease that is mainly transmitted by respiratory droplets and contact, and is generally infectious to human beings [1]. On January 11, 2020, after pathogenic nucleic acid testing, China reported 41 cases of pneumonia that were infected with the novel coronavirus [2]. On January 30, 2020, the World Health Organization (WHO) listed the novel coronavirus epidemic as a Public Health Emergency of International Concern [3]. As of February 20, 2020, a total of 75,465 confirmed cases and 2236 deaths have been reported in mainland China [4]. The novel coronavirus has caused great challenges and threats to public health in China and has quickly become the focus of worldwide attention. Information and discussions about COVID-19 have spread rapidly online, especially on social media.

To fight against COVID-19 and promote the prevention and control of the epidemic, researchers have recently made efforts in various aspects. The research has involved topics such as epidemiological research [5-7], diagnostic methods for COVID-19 [8-11], clinical characteristics of the disease [12-16], characteristics of disease transmission [17-19], development of candidate therapies [20-22], and the identification of animal hosts [22-26]. However, there has been no research conducted on the public's attention to COVID-19. Since public participation is required to prevent and control the epidemic spread of infectious diseases, it is extremely important to learn about the public's attention toward COVID-19 during the current epidemic. Such knowledge is of great significance when guiding people to respond appropriately to the epidemic and helping them learn how to cope with the sudden infectious diseases such as COVID-19, and it also supports social stability [27].

Social media has developed rapidly in recent years. Increasing numbers of public health departments and individuals are using social media platforms to communicate and share information during public health emergencies. Social media has become an important channel for promoting risk communication during the crisis [28,29]. The use of social media to measure public attention has also been gradually applied to research on infectious diseases such as H7N9 [27,30,31], Ebola [28,32-36], Zika virus [29,37,38], Middle East respiratory syndrome-related coronavirus [39], and Dengue fever [40]. In addition, to help countries respond better to the spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the WHO has issued operational planning guidelines, requiring countries to take specific actions around the eight pillars. Risk Communication and Community Engagement, one of the eight pillars, plays an important role in health emergencies and requires the promotion of active public communication to clear public confusion and misunderstanding [41]. Since the outbreak of the novel coronavirus, Sina Microblog, the main social media platform in China, has become increasingly important for the public to communicate about the COVID-19 epidemic. Therefore, it is of immediate importance for governments and nongovernmental organizations to study the public's attention toward the COVID-19 epidemic on Sina Microblog.

Sina Microblog ("Weibo" for short) is one of the most popular social media platforms in China and is the equivalent of Twitter in China. As of the fourth quarter of 2018, the number of monthly active users had reached 462 million, and approximately 200 million people are using Sina Microblog every day [42]. The Sina Microblog hot search list is the ranking of the most followed and hottest information on Sina Microblog, and is the most popular functional module in Sina Microblog apps [43,44]. This ranking is sorted according to the search volume of the hot topics that users searched for within a certain period. The higher the search volume is, the higher the ranking is, which directly reflects the public attention and attitude toward the topic.

This paper studies the public attention given to COVID-19 on Sina Microblog by searching for and analyzing topics related to the COVID-19 epidemic on the Sina Microblog hot search list from December 31, 2019, to February 20, 2020. We use a Chinese social media platform, Sina Weibo, to study the public attention of COVID-19 [45]. In this study, we describe the trend of public attention given to topics related to the COVID-19 epidemic and the hot topic keywords of public concern, analyze the emotional tendencies and trends of hot topics related to the COVID-19 epidemic, and conduct a visual cluster analysis of the hot topic content. This approach is used to obtain timely access to public responses so that the government and the health department can better communicate with the public on health issues and take appropriate measures to prevent and control the epidemic [35,39].

Methods

Research Overview

The research process mainly included five steps: (1) collecting topics related to the COVID-19 epidemic on the Sina Microblog hot search list; (2) segmenting the collected text into words; (3) describing the Sina Microblog search trend around the COVID-19 epidemic; (4) evaluating public opinion through word frequency and sentiment analysis; and (5) constructing a social network of public opinion through the subject analysis of the content. Each step is described in detail below.

Data Collection

We obtained information on COVID-19 in mainland China from the National Health Committee of the People's Republic of China [46]. The Wuhan Municipal Health Committee first reported viral pneumonia of unknown causes on December 31, 2019 [47]. This study collected topics on the Sina Microblog hot search list from December 31, 2019, to February 20, 2020, using Weibo Search Rank [48] from ENLIGHTENT [49], and selected topics and their search volume related to the COVID-19 epidemic. A total of 4056 topics related to the COVID-19 epidemic were on the hot search list, and 3234 remained after excluding duplicate topics. This group was used as the data basis for further processing, analysis, and discussion.

Data Processing

The topics related to the COVID-19 epidemic on the Sina Microblog hot search list from December 31, 2019, to February 20, 2020, were summarized and classified daily in chronological

order. We used the Chinese word segmentation function in ROST Content Mining System Version 6.0 (ROST CM6.0) to segment the content of hot search topics on Sina Microblog into words [50]. After word segmentation, the text was processed by merging synonyms and deleting nonsense words to provide a basis for subsequent research.

Data Analysis

Trend Analysis

We plotted the number of topics related to the COVID-19 epidemic on the Sina Microblog hot search list and the cumulative search volume by date to explore the public's attention toward the COVID-19 epidemic over time. A Pearson correlation analysis was used to explore the relationship between the number of hot search topics related to the COVID-19 epidemic and the cumulative search volume.

Word Frequency and Sentiment Analysis

After the Chinese word segmentation and invalid word filtering, we used ROST CM6.0 software to perform word frequency statistics and sentiment analysis on the Sina Microblog hot topics related to the COVID-19 epidemic. We calculated the frequency of keywords appearing in the hot search topics on Sina Microblog, explored the sentiment tendency of each hot search topic to the COVID-19 epidemic, and calculated the emotional score.

Social Network Analysis and Visualization

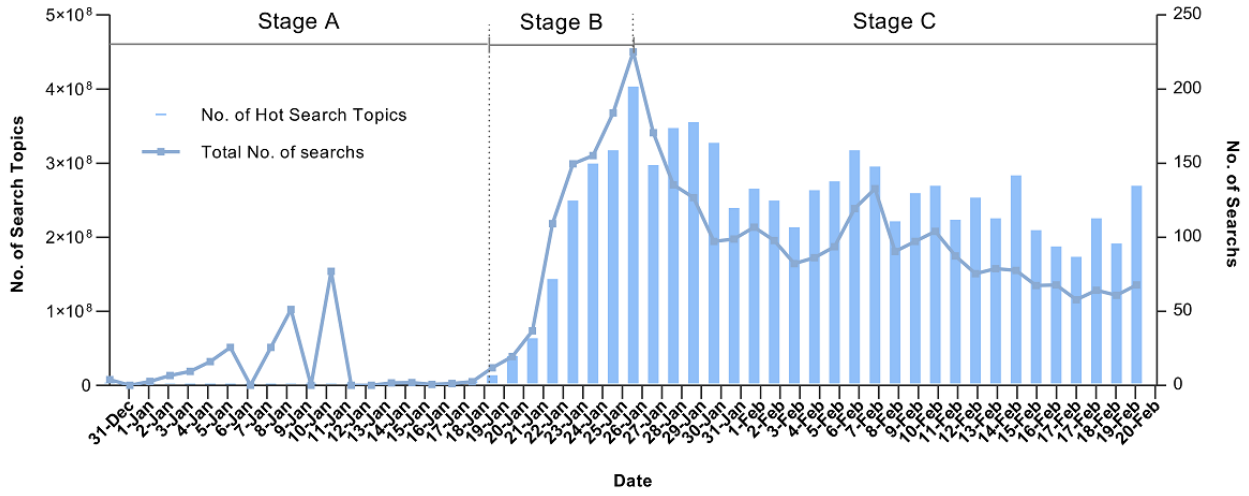
The high-frequency keywords and their frequencies were extracted from the Sina Microblog hot topic text after word segmentation and invalid word filtering. We used the VOSviewer software developed by Nees Jan van Eck and Ludo Waltman of the Centre for Science and Technology Studies of Leiden University in the Netherlands to create a visual knowledge map of keyword co-occurrence analysis and cluster analysis, using the keyword co-occurrence frequency as the weight [51].

Results

Search Trend of COVID-19 Epidemic on Sina Microblog Hot Search List

Figure 1 lists the number of COVID-19 epidemic-related topics and the cumulative search volume on the Sina Microblog hot search list from December 31, 2019, to February 20, 2020. The Pearson correlation analysis shows that the number of topics related to the COVID-19 epidemic is positively correlated with the cumulative search volume of the topic per day (r=0.767, P<.001). In other words, the more topics that are listed, the greater the cumulative search volume related to the COVID-19 epidemic per day. We can see that the public's attention to the COVID-19 epidemic on Sina Microblog can be divided into three stages. Stage A (December 31, 2019, to January 18, 2020) has low and unstable public attention, which represents the incubation period. Stage B (January 19-26, 2020) has a concentrated increase in public attention, which represents the epidemic period. Stage C (January 27 to February 20, 2020) demonstrates continued public attention to the epidemic, representing a widespread period.

Figure 1. The number of topics and the cumulative search volume of the coronavirus disease epidemic on the Sina Microblog hot search list.



Frequency and Distribution of Hot Search Keywords Related to the COVID-19 Epidemic on Sina Microblog

The top 15 keywords and their frequencies for the three stages of the public's attention to the COVID-19 epidemic are shown in Table 1. "Wuhan," "case," and "pneumonia" always appeared in the three periods as hot keywords, and the remaining keywords in the different periods were slightly different. In stage A, "unknown cause" and "novel coronavirus" were the main keywords, indicating that in the initial stage of the epidemic, viral pneumonia had just been discovered, and the cause was unknown. After the pathogen was initially identified as a novel coronavirus, the public began to search for information on the novel coronavirus to learn relevant knowledge. Compared with stage A, stage B had new keywords such as "new," "mask," "first case," and "first-level response." The reason for this is that the COVID-19 epidemic had spread across the country during this stage. The first cases had appeared successively throughout the country, and the number of confirmed cases was increasing. The outbreak of COVID-19

made the people and the government aware of the importance of prevention. People began to buy masks, and governments at all levels initiated first-level responses to major public health emergencies. Compared with the previous two stages, the main keywords appearing in stage C were “discharged,” “national,” “materials,” and “Huoshenshan Hospital.” At this stage, the public’s attention had shifted to material donation and medical service assistance in the key epidemic areas in Wuhan. Moreover, the epidemic had spread throughout the country, and the public was more concerned about the rehabilitation of patients.

Table 1. Top 15 keywords about the COVID-19 epidemic on the Sina Microblog hot search list in the three periods.

Ranking	Stage A (Dec 31, 2019, to Jan 18, 2020)		Stage B (Jan 19-26, 2020)		Stage C (Jan 27 to Feb 20, 2020)	
	Keyword	Frequency, n	Keyword	Frequency, n	Keyword	Frequency, n
1	Wuhan	17	COVID-19[a]	141	Wuhan	316
2	Pneumonia	14	Wuhan	124	COVID-19	283
3	Unknown cause	9	Confirmed diagnosis	87	Case	212
4	Novel coronavirus	6	Case	84	Confirmed diagnosis	212
5	Patient	5	Add	52	Add	199
6	Case	5	Masks	37	Epidemic	190
7	Thailand	4	Epidemic	31	Hubei	159
8	Leave hospital	2	Hubei	29	Masks	149
9	Death	2	Startup	25	Patient	149
10	Person-to-person spread	2	First case	22	Hospital	98
11	Add	2	Pneumonia	20	Leave hospital	90
12	Japan	2	Beijing	20	Nationwide	83
13	COVID-19	2	Patient	19	Beijing	72
14	Epidemic disease	2	First-level response	18	Goods and materials	59
15	Eliminate	2	Novel coronavirus	17	Huoshenshan hospital	59

[a]COVID-19: coronavirus disease.

Sentiment Analysis of the Hot Search Topics Related to the COVID-19 Epidemic on Sina Microblog

After segmentation, we imported the Sina Microblog hot topic text related to the COVID-19 epidemic into the ROST CM6.0 sentiment analysis tool and obtained the sentiment and proportion of the three stages of public attention given to the topics related to the COVID-19 epidemic (Figure 2). Emotions were classified as positive, negative, and neutral emotions. Based on this, positive and negative emotions were subdivided into three categories: general, moderate, and high. Neutral emotions were not subdivided. We found that the sentiment of the hot topics of the COVID-19 epidemic on Sina Microblog in stage A tended to be negative, accounting for 58% (n=14/24), of which high and moderate negative emotions accounted for 13% (n=3) and 17% (n=4), respectively, and positive emotions accounted for the lowest proportion at 8% (n=2). In stage B, the majority of the hot search topics were neutral; however, 27.2% (n=174/639) of the hot search topics showed negative emotions, with 0.3% (n=2) being highly negative and 3.6% (n=23) being moderately negative, while positive emotions accounted for 13.3% (n=85). In stage C, 18.01% (n=470/2610) of the hot search topics showed positive emotions, of which moderate positive emotions accounted for 2.26% (n=59) and highly positive emotions accounted for 0.31% (n=8); in addition, 28.24% (n=737) of the hot search topics showed negative emotions. Based on the comprehensive analysis of the three-stage emotional tendencies, the public’s negative emotions toward the COVID-19 epidemic were weakened as a whole, and their positive emotions were generally enhanced.

Figure 3 shows the trend for the proportion of the daily emotional tendencies of the hot search topics related to the COVID-19 epidemic from December 31, 2019, to February 20, 2020. We can see that the three kinds of emotions are relatively unstable before January 9, 2020. From January 9-20, 2020, negative emotions accounted for the largest proportion, followed by neutral emotions, and positive emotions accounted for the smallest proportion.

Compared with the emotions prior to January 20, 2020, the positive emotions of the hot search topics related to the COVID-19 epidemic after January 20, 2020, were generally on the rise; the negative emotions declined on the whole, and the emotions tended to be stable, as seen in Figure 3. This outcome shows that as the COVID-19 epidemic began to spread throughout the country after January 20, 2020, the public eased their concerns and fears caused by their uncertainty toward and ignorance of the epidemic, and responded to the epidemic with a more objective attitude.

Figure 2. The sentiment statistics of the Sina Microblog hot search topics related to the coronavirus disease epidemic.

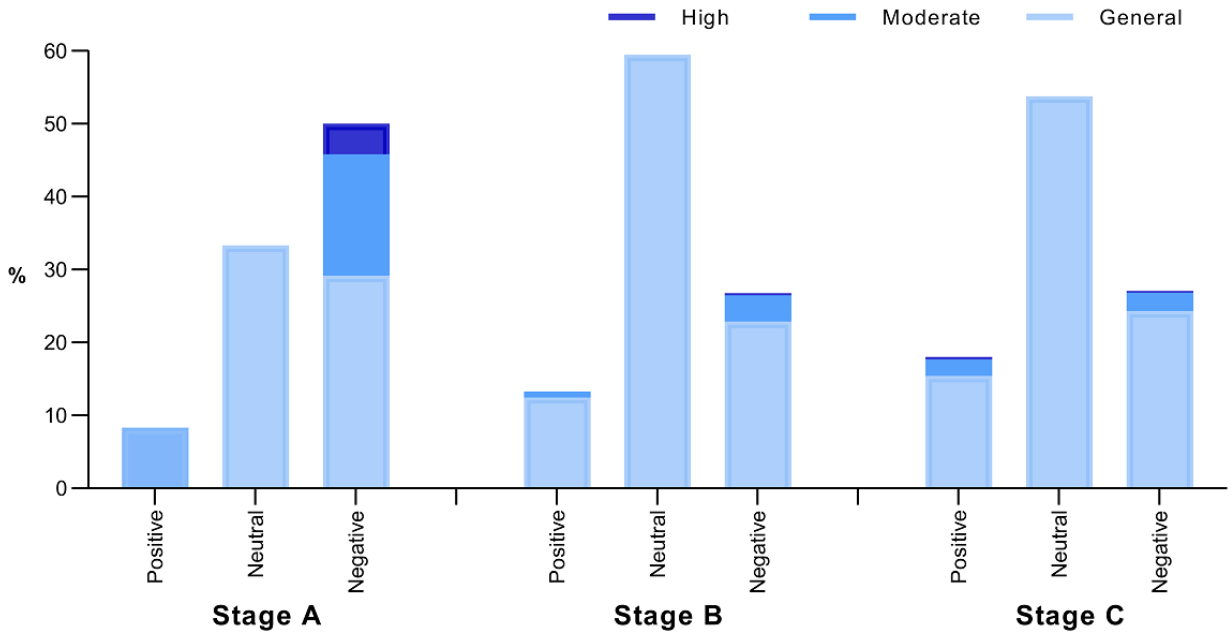
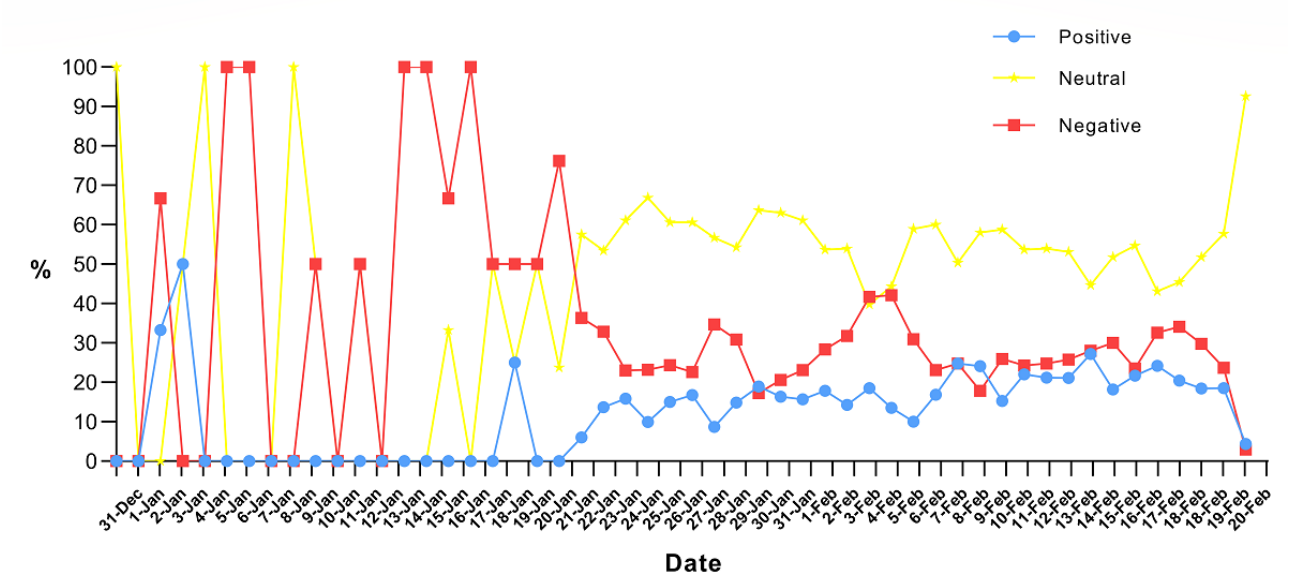


Figure 3. The sentiment trend of hot search topics related to the coronavirus disease epidemic from December 31, 2019, to February 20, 2020.



Social Semantic Network Analysis of Hot Search Topics Related to the COVID-19 Epidemic on Sina Microblog

To explore the themes reflected by the related topics of the COVID-19 epidemic on the Sina Microblog hot search list, this study used VOSviewer to generate clusters and co-occurrence networks of topic keywords. The results are shown in Figure 4. In the figure, the larger the size of the nodes and the font, the greater the weight of the keyword is, and it is closer to the core position. The connection between the nodes indicates that the keywords on the two nodes have appeared together. The thicker the connection is, the higher the frequency of co-occurrence and the closer the connection is between the nodes. Nodes with the same color belong to the same cluster.

According to the network visualization graph constructed in Figure 4, we can see that the keyword “COVID-19” is at the core node position, and the two nodes “Wuhan” and “add” are next. The core topic of public concern about the COVID-19 epidemic was COVID-19 itself, and the public was extremely concerned about the status of the epidemic in Wuhan and the new cases. We can divide the high-frequency keywords of topics into five clusters.

Cluster 1 (red cluster) involved discussion related to new cases, the outbreak of COVID-19 across the country, and the impact of the epidemic on the resumption of school and work. Examples of keywords included “case,” “newly increased,” “appear,” “Hubei Province,” and “Beijing” (Figure 5).

Figure 4. Social network of high-frequency keywords in the Sina Microblog hot search topics related to the COVID-19 epidemic.

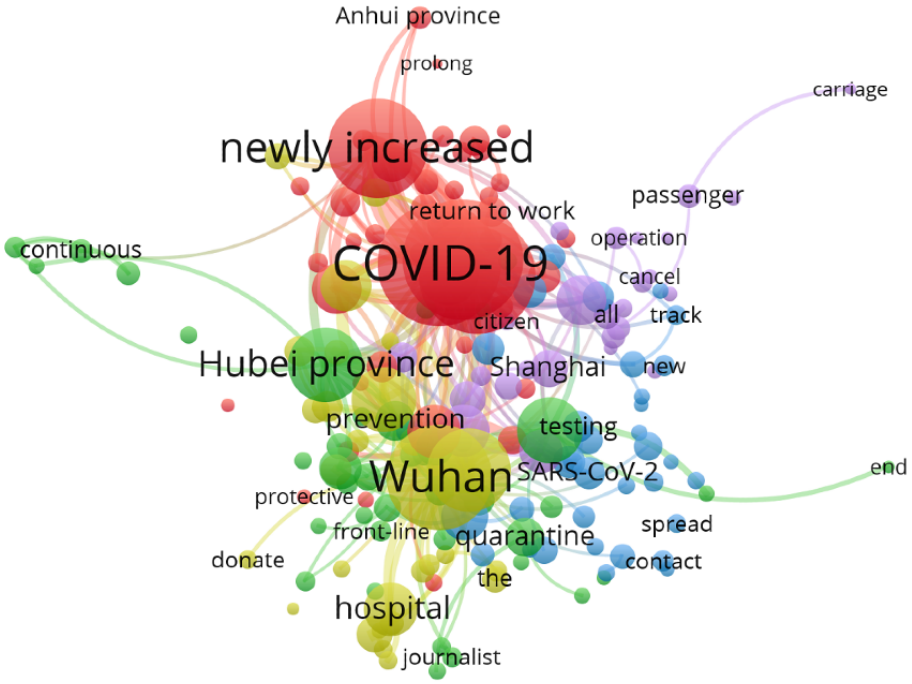
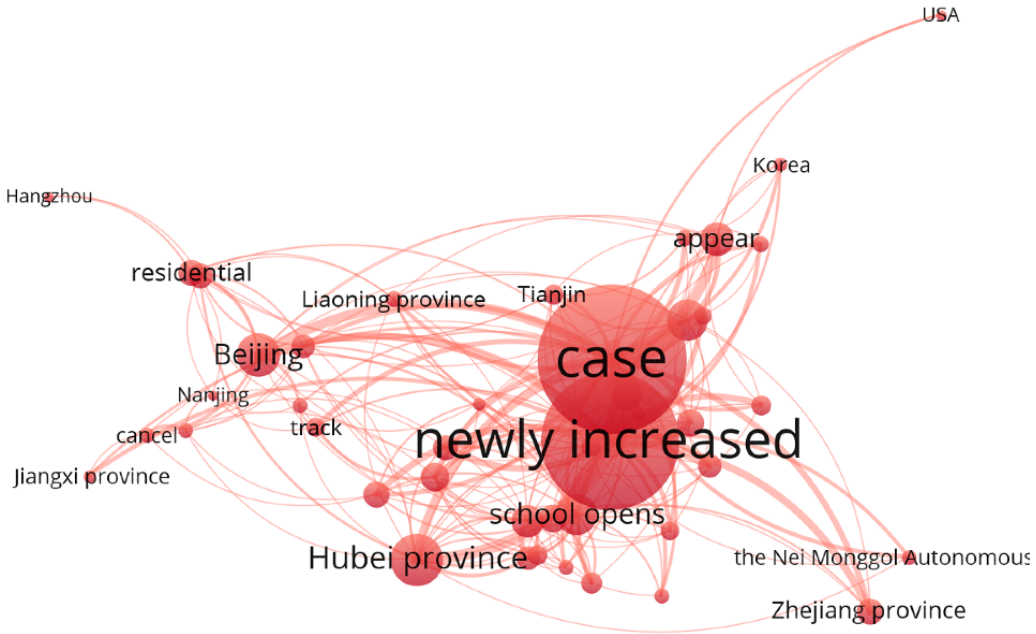


Figure 5. Cluster 1.



Cluster 2 (green cluster) involved searches for news reports on the frontline of the epidemic and related measures for prevention and control. Examples of keywords included “Wuhan,” “front-line,” “inpatient,” “quarantine,” and “living at home” (Figure 6).

Cluster 3 (blue cluster) involved searches for interpretations of the epidemic situation and prevention and control, and the knowledge of experts and relevant health departments, as well as the discussion on the source of the infection. The keywords in this cluster included “academician,” “WHO,” “face mask,” “epidemic prevention,” “symptom,” “virus,” and “SARS-CoV-2” (Figure 7).

Cluster 4 (yellow cluster) included searches for frontline medical services such as frontline hospital construction and medical team support. Examples of keywords in this cluster included “patient,” “hospital,” “Leishenshan,” “Huoshenshan,” and “first batch” (Figure 8).

Cluster 5 (purple cluster) included searches for the worldwide spread of the disease and searching online for fellow passengers with confirmed cases. Examples of keywords in this cluster included “passenger,” “carriage,” “cruise,” “suspend,” “infection,” and “Japan” (Figure 9).

Figure 6. Cluster 2.

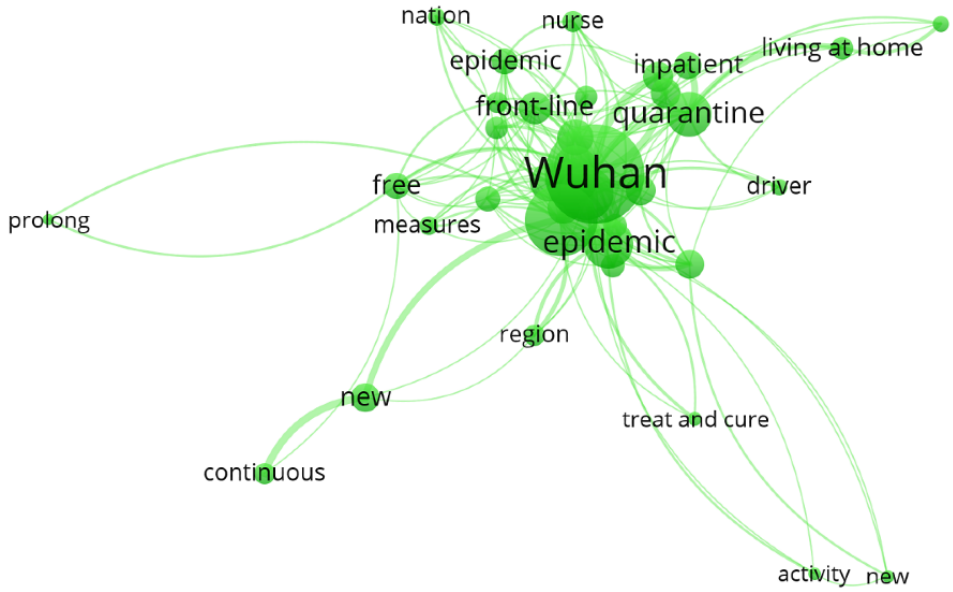


Figure 7. Cluster 3.

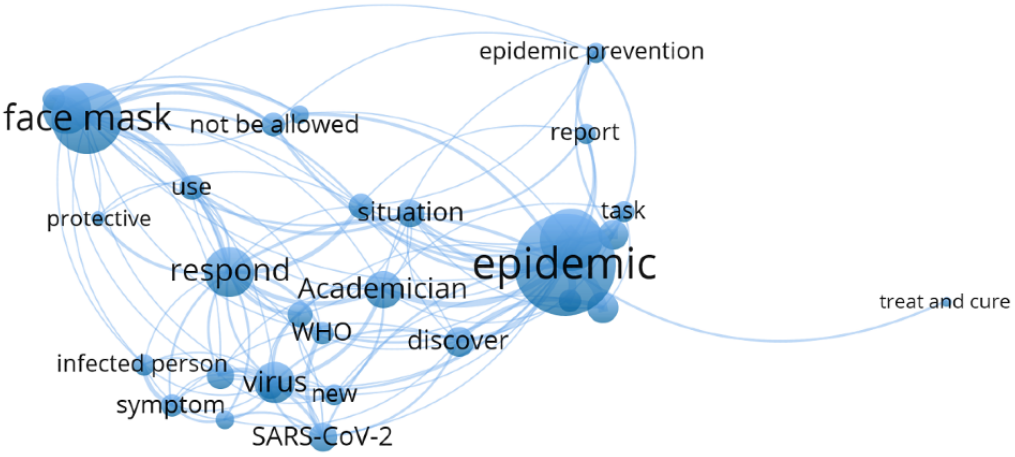


Figure 8. Cluster 4.

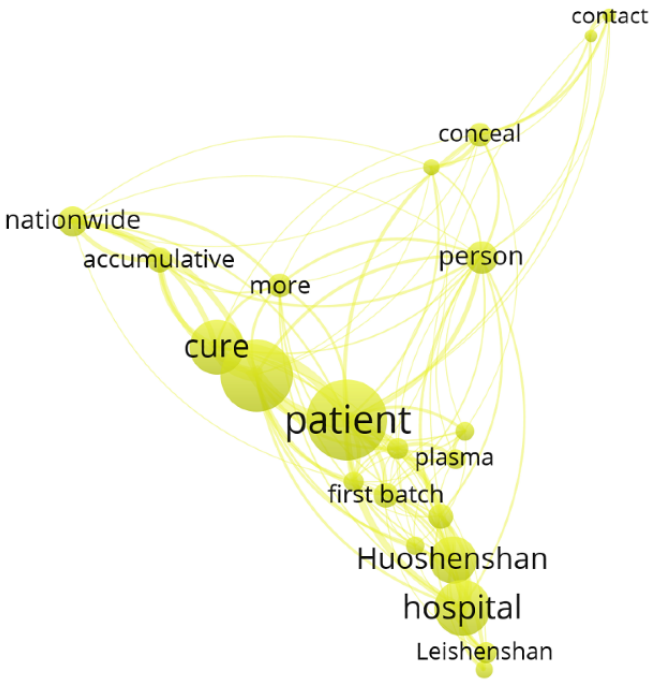


Figure 9. Cluster 5.

Discussion

Principal Results

In recent years, as people increasingly seek health information online, social media has played an increasingly important role in public health emergencies [29,35]. However, there have been no relevant studies that have incorporated social media analysis into the public response to the COVID-19 epidemic. We used the Sina Microblog hot search list to analyze the public's attention toward the COVID-19 epidemic from December 31, 2019, to February 20, 2020, which was 52 days after the first disclosure of COVID-19 by the Chinese health department. There were four main findings from this study. First, we analyzed the changing trend of public attention given to the COVID-19 epidemic, which can be divided into three stages. Second, the hot topic keywords of the public's attention at each stage were slightly different. Third, the public's emotional tendency toward the hot search topics related to the COVID-19 epidemic changed from negative to neutral across the study period. As a whole, negative emotions weakened, and positive emotions increased. Fourth, we divided the topics of public concern about the COVID-19 epidemic into five categories through social semantic network analysis. This study analyzes

the public's response and attention given to the COVID-19 epidemic, which will help public health professionals monitor public response, identify public needs as early as possible, make timely public health prevention and control measures, and disseminate knowledge to citizens in a targeted manner to better respond to the current COVID-19 epidemic.

From December 31, 2019, to February 20, 2020, the public attention given to the COVID-19 epidemic on Sina Microblog can be divided into three stages. In the beginning, there was little public attention paid to the epidemic, and then the concentration of attention increased. Next, the public attention given to the epidemic generally declined, but people continued to pay attention. In the first stage (December 31, 2019, to January 18, 2020), the Chinese public paid less attention, as COVID-19 was only reported in Wuhan, China. On January 19, 2020, suspected cases appeared in Shanghai and Shenzhen, and body temperatures began to be measured at the Wuhan Airport and Railway Station. People began to notice the severity of the epidemic, and the level of attention paid to the COVID-19 epidemic began to increase until January 26, 2020. We can see that when the epidemic began to spread across the country, the public responded quickly to news about the COVID-19 outbreak on Sina Microblog [39]. This is because COVID-19 is a new infectious disease, which means that no effective treatment has been found and no corresponding vaccine has been developed. When the outbreak began, the public was eager to search for relevant knowledge and information online to meet their own protection needs [27,30]. In the third stage (January 27 to February 20, 2020), although the number of cases of COVID-19 were still increasing, the number of Sina Microblog topics and their search volume were decreasing. The reason for this may be because the epidemic information was relatively saturated, and it was difficult for people to acquire more new knowledge through a Sina Microblog search [27]. It may also be that people were no longer paying close attention to the epidemic dynamic to gain a sense of security as time went on, and the public's consciousness tends to be rational [52].

The keywords for the three different stages of the public's attention to the COVID-19 epidemic were slightly different. From the hot search keywords of each stage, we found that the public did not know much about the virus and its causes in the first stage. The main keywords of the search to seek relevant knowledge were "unknown cause" and "novel coronavirus." During the second stage, the epidemic began to spread throughout the country, the first cases were reported in different areas successively, and the number of confirmed cases continued to increase, thereby making the public and government aware of the importance of prevention. During the third stage, the epidemic spread widely throughout the country; Wuhan was the most seriously affected area, and the city was locked down. The public's attention was mainly shifted to material donation and medical assistance in Wuhan. Compared with qualitative research, it is more accurate to obtain the hot spots of public attention through keyword frequency analysis, as this process has a higher accuracy and more credibility in the research of hot spots and their development trends [40]. Based on our high-frequency keyword analysis of the Sina Microblog topics related to the COVID-19 epidemic, we can obtain information about the concerns and opinions of Sina Microblog users at different stages [53]. Studies have found that people's interest in infectious diseases on social media was linked to the latest news and major events. Studies have also shown that people will pay attention to and search for disease-related words as the spread of infectious diseases change [30].

According to the sentiment tendencies of the hot search topics related to the COVID-19 epidemic on Sina Microblog, the first stage of emotion was negative, and the second and third stages were neutral. On the whole, negative emotions weakened, and positive emotions increased. Previous studies have pointed out that there is also an important relationship between emotions and content on social media [54]. The content analysis of social networks has identified people's attitudes or reactions to specific health hazard events [30,39]. In the first stage, there was less public attention given to the Sina Microblog hot search list. Most of the topics related to the COVID-19 epidemic were about the notification of pneumonia and the virus, and the emotions tended to be negative. At that time, the public had a strong demand for information on public health emergencies such as infectious diseases. When the information demand could not be fully satisfied, the users' emotions were negative [40]. As the epidemic progressed to the second and third stages, the public sentiment tended to be neutral because increasingly more news was being reported at this stage, and objective events became the mainstream information on the Sina Microblog hot search list. The public reduced their previous levels of worries and fears about the epidemic; their negative feelings weakened, and their positive emotions increased. More hot search topics mentioned information about prevention or protection, which is conducive to public health communication and promotion.

We divided the COVID-19 topics with the highest levels of public concern into five categories: the situation of new cases of COVID-19 and its impact, frontline reporting of the epidemic and the measures of prevention and control, expert interpretation and discussion on the source of infection, medical services on the frontline of the epidemic, and focus on the worldwide epidemic and looking for suspected cases. From the search of the subject content, we can see that during the outbreak of the novel coronavirus, the public, the news media, and the health department all actively used Sina Microblog as a platform for disseminating information related to COVID-19 [29], indicating that Sina Microblog is a communication channel for both individuals and organizations to publicize COVID-19 symptoms, preventive measures, and related policies [37]. In addition, as a real time and extensive online platform, Sina Microblog provides a channel for information dissemination. For example, in this emergency outbreak, the public has made good use of the platform to find fellow passengers with confirmed cases, thereby playing an important role in preventing and controlling the disease transmission.

Limitations

There are some limitations to this study. First, our study was limited to the period 52 days after COVID-19 was first disclosed by the Chinese health department. Thus, the situation after February 20, 2020, was not included in this study. Second, we only used Sina Microblog as the social media platform in this

study, the data source for this study is relatively narrow. Considering that China is seriously aging and that the majority of internet users are young people, the age of Sina Microblog users is mainly concentrated in those 18-41 years [42], which cannot represent the population of all ages in China. Therefore, selection bias is inevitable in this study. On the other hand, the study excludes other popular social media data sources such as WeChat and ByteDance and is limited to the publicly available data on the Sina Microblog hot search list. Moreover, due to the lack of detailed information about the users who contributed to the search volume of the Sina Microblog hot search list, we could not describe the social demographic information of the Sina Microblog users, and we failed to obtain the geospatial distribution of Sina Microblog active users. Thus, we could not calculate the average attention of the public in different regions of China.

Future Work

Social media has played a vital role during the spread of the COVID-19 epidemic and has become the main way for the public to obtain epidemic information. However, the excessive amount of information about the novel coronavirus (some right, some wrong) on social media may make it difficult for people to find reliable sources and guidance when needed, and may even be harmful to people's health, leading to the outbreak of an infodemic [55]. Therefore, misinformation can be easily propagated and spread on social media. Apart from traditional "top-down" public health communication, provider-managed communication platforms such as WebMD (America), AskDr (Singapore), and HaoDF (China) can be used as an additional countermeasure available to develop a ground-up response to this issue. HaoDF [56], one of the largest doctor-patient communication platforms in China, launched an online free consultation service for COVID-19 during this epidemic. Doctors answered questions online for free 24 hours a day to help the public make a preliminary judgment about the disease and avoid blind medical treatment caused by panic. In addition, HaoDF summarized the common questions and answers of COVID-19 to let the public know more about the epidemic. These are the forms of responses, as potential solutions to avoid an infodemic, that governments can engage with, researchers can study, and innovators can develop in the future.

Conclusions

Our study found that social media platforms such as Sina Microblog can be used to measure public attention given to public health emergencies. Our study shows that a large amount of information about the COVID-19 epidemic was disseminated and received widespread public attention on Sina Microblog during the novel coronavirus epidemic. We have learned about the hotspots of public concern regarding the COVID-19 epidemic on Sina Microblog. These findings can help the government and health departments better communicate with the public about public health and then translate public health needs into practice to create targeted measures to prevent and control the spread of COVID-19.

Conflicts of Interest

None declared.

References

1. Special Expert Group for Control of the Epidemic of Novel Coronavirus Pneumonia of the Chinese Preventive Medicine Association. [An update on the epidemiological characteristics of novel coronavirus pneumonia（COVID-19)]. Zhonghua Liu Xing Bing Xue Za Zhi 2020 Feb 14;41(2):139-144. [doi: 10.3760/cma.j.issn.0254-6450.2020.02.002] [Medline: 32057211]
2. Wuhan Municipal Health Committee. Report of Wuhan Municipal Health Committee on viral pneumonia of unknown cause [in Chinese] URL: http://wjw.wuhan.gov.cn/front/web/showDetail/2020011109035 [accessed 2020-02-21]
3. World Health Organization. 2020 Jan 31. Novel Coronavirus (2019-nCoV) Situation Report - 11 URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200131-sitrep-11-ncov.pdf?sfvrsn=de7c0f7_4 [accessed 2020-02-21]
4. National Health Commission of the people's Republic of China. Pneumonia situation in novel coronavirus infection on January 21st [in Chinese] URL: http://www.nhc.gov.cn/xcs/yqtb/202001/930c021cdd1f46dc832fc27e0cc465c8.shtml [accessed 2020-02-21]
5. Sun K, Chen J, Viboud C. Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study. The Lancet Digital Health 2020 Apr;2(4):e201-e208. [doi: 10.1016/s2589-7500(20)30026-1]
6. Xu B, Kraemer MUG, Xu B, Gutierrez B, Mekaru S, Sewalk K, et al. Open access epidemiological data from the COVID-19 outbreak. The Lancet Infectious Diseases 2020 Feb:e. [doi: 10.1016/s1473-3099(20)30119-5]
7. Lipsitch M, Swerdlow DL, Finelli L. Defining the epidemiology of Covid-19 — studies needed. N Engl J Med 2020 Mar 26;382(13):1194-1196. [doi: 10.1056/nejmp2002125]
8. El-Tholoth M, Bau HH, Song J. A single and two-stage, closed-tube, molecular test for the 2019 Novel Coronavirus (COVID-19) at home, clinic, and points of entry. ChemRxiv 2020:e. [doi: 10.26434/chemrxiv.11860137]
9. Li Q, Ding X, Xia G, Geng Z, Chen F, Wang L, et al. A simple laboratory parameter facilitates early identification of COVID-19 patients. medRxiv 2020:e. [doi: 10.1101/2020.02.13.20022830]
10. To K, Tsang O, Chik-Yan Yip C, Chan KH, Wu TC, Chan JMC, et al. Consistent detection of 2019 novel coronavirus in saliva. Clin Infect Dis 2020 Feb 12:e [FREE Full text] [doi: 10.1093/cid/ciaa149] [Medline: 32047895]

11. Yang Y, Yang M, Shen C, Wang F, Yuan J, Li J, et al. Evaluating the accuracy of different respiratory specimens in the laboratory diagnosis and monitoring the viral shedding of 2019-nCoV infections. medRxiv 2020:e. [doi: 10.1101/2020.02.11.20021493]
12. Pan X, Chen D, Xia Y, Wu X, Li T, Ou X, et al. Asymptomatic cases in a family cluster with SARS-CoV-2 infection. The Lancet Infectious Diseases 2020 Apr;20(4):410-411. [doi: 10.1016/s1473-3099(20)30114-6]
13. Xu X, Wu X, Jiang X, Xu K, Ying L, Ma C, et al. Clinical findings in a group of patients infected with the 2019 novel coronavirus (SARS-Cov-2) outside of Wuhan, China: retrospective case series. BMJ 2020 Feb 19;368:m606. [doi: 10.1136/bmj.m606] [Medline: 32075786]
14. Silverstein WK, Stroud L, Cleghorn GE, Leis JA. First imported case of 2019 novel coronavirus in Canada, presenting as mild pneumonia. The Lancet 2020 Feb;395(10225):734. [doi: 10.1016/s0140-6736(20)30370-6]
15. Pan F, Ye T, Sun P, Gui S, Liang B, Li L, et al. Time course of lung changes on chest CT during recovery from 2019 novel coronavirus (COVID-19) pneumonia. Radiology 2020 Feb 13:e. [doi: 10.1148/radiol.2020200370]
16. Fang Y, Zhang H, Xu Y, Xie J, Pang P, Ji W. CT manifestations of two cases of 2019 novel coronavirus (2019-nCoV) pneumonia. Radiology 2020 Apr;295(1):208-209. [doi: 10.1148/radiol.2020200280] [Medline: 32031481]
17. Liang W, Guan W, Chen R, Wang W, Li J, Xu K, et al. Cancer patients in SARS-CoV-2 infection: a nationwide analysis in China. The Lancet Oncology 2020 Mar;21(3):335-337. [doi: 10.1016/s1470-2045(20)30096-6]
18. Chen J. Pathogenicity and transmissibility of 2019-nCoV-a quick overview and comparison with other emerging viruses. Microbes Infect 2020 Mar;22(2):69-71. [doi: 10.1016/j.micinf.2020.01.004] [Medline: 32032682]
19. Mizumoto K, Kagaya K, Chowell G. Early epidemiological assessment of the transmission potential and virulence of 2019 Novel Coronavirus in Wuhan City: China, 2019-2020. medRxiv 2020:e. [doi: 10.1101/2020.02.12.20022434]
20. Micholas S, Jeremy C. Repurposing therapeutics for COVID-19: supercomputer-based docking to the SARS-CoV-2 viral spike protein and viral spike protein-human ACE2 interface. ChemRxiv 2020:e. [doi: 10.26434/chemrxiv.11871402.v3]
21. Anh-Tien T, Francesco G, Michael H, Fuqiang B, Artem C. Rapid identification of potential inhibitors of SARS-CoV-2 main protease by deep docking of 1.3 billion compounds. ChemRxiv 2020:e. [doi: 10.26434/chemrxiv.11860077]
22. Rimanshee A, Amit D, Vishal P, Mukesh K. Potential inhibitors against papain-like protease of novel coronavirus (SARS-CoV-2) from FDA approved drugs. ChemRxiv 2020:e. [doi: 10.26434/chemrxiv.11860011]
23. Liu P, Jiang JZ, Wan XF, Hua Y, Wang X, Hou F, et al. Are pangolins the intermediate host of the 2019 novel coronavirus (2019-nCoV)? bioRxiv 2020:e. [doi: 10.1101/2020.02.18.954628]
24. Xiao K, Zhai J, Feng Y, Zhou N, Zhang X, Zou JJ, et al. Isolation and characterization of 2019-nCoV-like coronavirus from Malayan pangolins. bioRxiv 2020:e. [doi: 10.1101/2020.02.17.951335]
25. Lam T, Shum M, Zhu H, Tong YG, Ni XB, Liao YS, et al. Identification of 2019-nCoV related coronaviruses in Malayan pangolins in southern China. bioRxiv 2020:e. [doi: 10.1101/2020.02.13.945485]
26. Wong M, Javornik CS, Ajami N, Petrosino J. Evidence of recombination in coronaviruses implicating pangolin origins of nCoV-2019. bioRxiv 2020:e. [doi: 10.1101/2020.02.07.939207]
27. Gu H, Chen B, Zhu H, Jiang T, Wang X, Chen L, et al. Importance of Internet surveillance in public health emergency control and prevention: evidence from a digital epidemiologic study during avian influenza A H7N9 outbreaks. J Med Internet Res 2014 Jan 17;16(1):e20 [FREE Full text] [doi: 10.2196/jmir.2911] [Medline: 24440770]
28. Househ M. Communicating Ebola through social media and electronic news media outlets: a cross-sectional study. Health Informatics J 2016 Sep;22(3):470-478. [doi: 10.1177/1460458214568037] [Medline: 25656678]
29. Gui X, Wang Y, Kou Y, Reynolds TL, Chen Y, Mei Q, et al. Understanding the patterns of health information dissemination on social media during the Zika outbreak. AMIA Annu Symp Proc 2017;2017:820-829 [FREE Full text] [Medline: 29854148]
30. Chen Y, Zhang Y, Xu Z, Wang X, Lu J, Hu W. Avian Influenza A (H7N9) and related Internet search query data in China. Sci Rep 2019 Jul 18;9(1):10434. [doi: 10.1038/s41598-019-46898-y] [Medline: 31320681]
31. Xie T, Yang Z, Yang S, Wu N, Li L. Correlation between reported human infection with avian influenza A H7N9 virus and cyber user awareness: what can we learn from digital epidemiology? Int J Infect Dis 2014 May;22:1-3. [doi: 10.1016/j.ijid.2013.11.013] [Medline: 24576842]
32. Fung IC, Fu K, Chan C, Chan BSB, Cheung C, Abraham T, et al. Social media's initial reaction to information and misinformation on Ebola, August 2014: facts and rumors. Public Health Rep 2016;131(3):461-473 [FREE Full text] [doi: 10.1177/003335491613100312] [Medline: 27252566]
33. Fung IC, Duke CH, Finch KC, Snook KR, Tseng P, Hernandez AC, et al. Ebola virus disease and social media: a systematic review. Am J Infect Control 2016 Dec 01;44(12):1660-1671. [doi: 10.1016/j.ajic.2016.05.011] [Medline: 27425009]
34. Wong R, Harris JK. Geospatial distribution of local health department tweets and online searches about Ebola during the 2014 Ebola outbreak. Disaster Med Public Health Prep 2018 Jun;12(3):287-290. [doi: 10.1017/dmp.2017.69] [Medline: 28835312]
35. Feng S, Hossain L, Crawford JW, Bossomaier T. Quantifying Network Dynamics and Information Flow Across Chinese Social Media During the African Ebola Outbreak. Disaster med. public health prep 2017 Aug 01;12(1):26-37. [doi: 10.1017/dmp.2017.29]
36. Seltzer E, Jean N, Kramer-Golinkoff E, Asch D, Merchant R. The content of social media's shared images about Ebola: a retrospective study. Public Health 2015 Sep;129(9):1273-1277. [doi: 10.1016/j.puhe.2015.07.025] [Medline: 26285825]

37. Seltzer E, Horst-Martz E, Lu M, Merchant R. Public sentiment and discourse about Zika virus on Instagram. Public Health 2017 Sep;150:170-175. [doi: 10.1016/j.puhe.2017.07.015] [Medline: 28806618]
38. Fu K, Liang H, Saroha N, Tse ZTH, Ip P, Fung IC. How people react to Zika virus outbreaks on Twitter? A computational content analysis. Am J Infect Control 2016 Dec 01;44(12):1700-1702. [doi: 10.1016/j.ajic.2016.04.253] [Medline: 27566874]
39. Fung IC, Fu K, Ying Y, Schaible B, Hao Y, Chan C, et al. Chinese social media reaction to the MERS-CoV and avian influenza A(H7N9) outbreaks. Infect Dis Poverty 2013 Dec 20;2(1):31. [doi: 10.1186/2049-9957-2-31] [Medline: 24359669]
40. Wang XY, Hao YH, Wu QH, Xu W, Wang XF, Dai YJ, et al. Visual analysis of official microblogging network public opinion on Dengue events [in Chinese]. Chinese Preventive Medicine 2018;19(06):401-406. [doi: 10.16506/j.1009-6639.2018.06.001]
41. World Health Organization. Coronavirus disease (COVID-19) technical guidance: risk communication and community engagement URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/risk-communication-and-community-engagement [accessed 2020-04-13]
42. Sina microblog data center. 2018 Sina microblog user development report [in Chinese] URL: https://data.weibo.com/report/reportDetail?id=433 [accessed 2020-02-21]
43. Chen Y. Analysis of the social functions of microblog hot search in the perspective of development communication [in Chinese]. Research on Transmission Competence 2019;3(04):76.
44. Zhang Y. Analysis of the reasonability of microblog hot search list in the perspective of communication [in Chinese]. Journal of News Research 2018;9(16):29-31.
45. Hu G, Han X, Zhou H, Liu Y. Public perception on healthcare services: evidence from social media platforms in China. Int J Environ Res Public Health 2019 Apr 10;16(7):e. [doi: 10.3390/ijerph16071273] [Medline: 30974729]
46. National Health Commission of the People's Republic of China. Efforts to prevent and control epidemic situation of COVID-19 [in Chinese] URL: http://www.nhc.gov.cn/xcs/yqtb/list_gzbd.shtml [accessed 2020-02-21]
47. Wuhan Municipal Health Committee. Report of Wuhan Municipal Health Committee on the current situation of pneumonia in our city [in Chinese] URL: http://wjw.wuhan.gov.cn/front/web/showDetail/2019123108989 [accessed 2020-02-21]
48. Weibo Search Rank [in Chinese]. URL: https://www.enlightent.cn/research/rank/weiboSearchRank [accessed 2020-02-21]
49. ENLIGHTENT [in Chinese]. URL: https://www.enlightent.cn/ [accessed 2020-02-21]
50. Fang Q. Research on the impact of ROST Content Mining System on content analysis [in Chinese]. Course Education Research 2014(01):234-235. [doi: 10.1093/oxfordhb/9780199811755.013.008]
51. VOSviewer - Visualizing scientific landscapes. URL: https://www.vosviewer.com/ [accessed 2020-02-21]
52. Liu B, Wang Z, Qi X, Zhang X, Chen H. Assessing cyber-user awareness of an emerging infectious disease: evidence from human infections with avian influenza A H7N9 in Zhejiang, China. Int J Infect Dis 2015 Nov;40:34-36. [doi: 10.1016/j.ijid.2015.09.017] [Medline: 26432409]
53. Lazard AJ, Wilcox GB, Tuttle HM, Glowacki EM, Pikowski J. Public reactions to e-cigarette regulations on Twitter: a text mining analysis. Tob Control 2017 Dec;26(e2):e112-e116. [doi: 10.1136/tobaccocontrol-2016-053295] [Medline: 28341768]
54. Massey PM, Leader A, Yom-Tov E, Budenz A, Fisher K, Klassen AC. Applying multiple data collection tools to quantify human papillomavirus vaccine communication on Twitter. J Med Internet Res 2016 Dec 05;18(12):e318 [FREE Full text] [doi: 10.2196/jmir.6670] [Medline: 27919863]
55. Zarocostas J. How to fight an infodemic. The Lancet 2020 Feb;395(10225):676. [doi: 10.1016/s0140-6736(20)30461-x]
56. HaoDF [in Chinese]. URL: https://www.haodf.com/ [accessed 2020-04-13]

Abbreviations

COVID-19: coronavirus disease
ROST CM6.0: ROST Content Mining System version 6.0
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
WHO: World Health Organization

Edited by G Eysenbach; submitted 23.03.20; peer-reviewed by D Gunasekeran, Y Liu; comments to author 09.04.20; revised version received 13.04.20; accepted 20.04.20; published 04.05.20.

Please cite as:
Zhao Y, Cheng S, Yu X, Xu H
Chinese Public's Attention to the COVID-19 Epidemic on Social Media: Observational Descriptive Study
J Med Internet Res 2020;22(5):e18825
URL: https://www.jmir.org/2020/5/e18825
doi:10.2196/18825
PMID:32314976

Original Paper

Conversations and Medical News Frames on Twitter: Infodemiological Study on COVID-19 in South Korea

Han Woo Park[1,2], PhD; Sejung Park[3], PhD; Miyoung Chong[4], MA

[1]Department of Media & Communication, Interdisciplinary Graduate Programs of Digital Convergence Business and East Asian Cultural Studies, Yeungnam University, Gyeongsan-si, Republic of Korea

[2]Cyber Emotions Research Institute, Gyeongsan-si, Republic of Korea

[3]Tim Russert Department of Communication, John Carroll University, Cleveland Heights, OH, United States

[4]College of Information, University of North Texas, Denton, TX, United States

Corresponding Author:
Sejung Park, PhD
Tim Russert Department of Communication
John Carroll University
1 John Carroll Blvd, University Heights
Cleveland Heights, OH, 44118
United States
Phone: 1 216 397 4722
Email: sjpark@jcu.edu

Abstract

Background: SARS-CoV-2 (severe acute respiratory coronavirus 2) was spreading rapidly in South Korea at the end of February 2020 following its initial outbreak in China, making Korea the new center of global attention. The role of social media amid the current coronavirus disease (COVID-19) pandemic has often been criticized, but little systematic research has been conducted on this issue. Social media functions as a convenient source of information in pandemic situations.

Objective: Few infodemiology studies have applied network analysis in conjunction with content analysis. This study investigates information transmission networks and news-sharing behaviors regarding COVID-19 on Twitter in Korea. The real time aggregation of social media data can serve as a starting point for designing strategic messages for health campaigns and establishing an effective communication system during this outbreak.

Methods: Korean COVID-19-related Twitter data were collected on February 29, 2020. Our final sample comprised of 43,832 users and 78,233 relationships on Twitter. We generated four networks in terms of key issues regarding COVID-19 in Korea. This study comparatively investigates how COVID-19-related issues have circulated on Twitter through network analysis. Next, we classified top news channels shared via tweets. Lastly, we conducted a content analysis of news frames used in the top-shared sources.

Results: The network analysis suggests that the spread of information was faster in the Coronavirus network than in the other networks (Corona19, Shincheon, and Daegu). People who used the word “Coronavirus” communicated more frequently with each other. The spread of information was faster, and the diameter value was lower than for those who used other terms. Many of the news items highlighted the positive roles being played by individuals and groups, directing readers’ attention to the crisis. Ethical issues such as deviant behavior among the population and an entertainment frame highlighting celebrity donations also emerged often. There was a significant difference in the use of nonportal (n=14) and portal news (n=26) sites between the four network types. The news frames used in the top sources were similar across the networks (P=.89, 95% CI 0.004-0.006). Tweets containing medically framed news articles (mean 7.571, SD 1.988) were found to be more popular than tweets that included news articles adopting nonmedical frames (mean 5.060, SD 2.904; N=40, P=.03, 95% CI 0.169-4.852).

Conclusions: Most of the popular news on Twitter had nonmedical frames. Nevertheless, the spillover effect of the news articles that delivered medical information about COVID-19 was greater than that of news with nonmedical frames. Social media network analytics cannot replace the work of public health officials; however, monitoring public conversations and media news that propagates rapidly can assist public health professionals in their complex and fast-paced decision-making processes.

(J Med Internet Res 2020;22(5):e18897) doi:10.2196/18897

KEYWORDS

infodemiology; COVID-19; SARS-CoV-2; coronavirus; Twitter; South Korea; medical news; social media; pandemic; outbreak; infectious disease; public health

Introduction

Background

SARS-CoV-2 (severe acute respiratory coronavirus 2) is spreading rapidly around the world, and the number of associated deaths has also been increasing. At the end of February 2020, the virus was spreading in South Korea following its initial outbreak in China, making Korea the new center of global attention. Mass infection occurred in Korea due to a closed religious group called *Shincheonji* in the greater Daegu metropolitan city, the fourth largest city in Korea [1,2].

Social media has been criticized often amid the current coronavirus disease (COVID-19) pandemic, mainly due to their use as a medium for the quick spread of fake news [3,4], but no systematic research has yet been conducted on this issue. Social media functions as a convenient source of information in dangerous situations [5]. Since the creation of social networking services (SNS) such as Facebook, Twitter, and YouTube, the speed of information transmission in disaster contexts has accelerated across social, cultural, and geographical boundaries. Real time information exchange through various SNS can facilitate the wider diffusion of risk information not only for "friends" but also for wider communities.

Using an infodemiological approach, this study analyzes networking trends in public conversations and news-sharing behavior regarding COVID-19, particularly in Daegu, South Korea, on Twitter. The Pew Research Center reported that approximately 75% of Twitter users visit Twitter.com to read the news [6]. Allowed up to 280 characters, Twitter users share thoughts and emotions through "tweets" and "retweets," which creates conversational and networked relationships on Twitter. The pattern of interactions between Twitter users may vary according to their interests and engagement with COVID-19. This study examines conversations on Twitter in relation to the greater Daegu metropolitan city cluster that was closely related to the members of the Shincheonji group, who contributed significantly to spreading the virus in the area. Four Twitter networks were chosen—Coronavirus, Corona19, Daegu, and Shincheonji—to represent the major issues regarding the COVID-19 crisis in the greater Daegu area. The keyword "Corona19" was included instead of "COVID-19" (coronavirus disease) because "Corona19" was announced as the official term for COVID-19 in Korea.

The Twitterverse examined in this context includes diverse messages on topics such as nationwide emergency relief efforts, media news, mass condolences, requests for central and regional governmental measures, and the provision of crucial medical information. The fact that these four networks have similar network sizes allows their conversational patterns and news diffusion to be easily compared.

Overcoming the current COVID-19 crisis may require increasingly diverse forms of data and more complex models. Handling the real time aggregation and artificial intelligence-based analytics of social media, media news, academic publications, and other data sets is a daunting task. Nevertheless, this study could serve as a starting point for designing strategic messages for health campaigns and establishing an effective communication channel system.

Infodemiology

Infodemiology is a growing area of research that aims to inform public health officials and develop public policies using informatics for the analysis of health data produced and consumed online [7]. The advantage of infodemiology is its capacity to obtain real time health-related data from unstructured, textual, image, or user-generated information communicated via electronic media such as websites, blogs, and social network sites [8].

Infodemiology studies have covered a wide range of topics. These include information search behaviors such as Ebola- or vaccination-related information [9,10], health-related news coverage [11], public health issues and awareness of diseases after the death of a celebrity [10], disparities in health information access and availability [12], public discussion and information sharing [13], and government risk communication strategies [14].

The trustworthiness of user-created information is questionable [15]. However, recent studies suggest that publicly available social media data such as those on Twitter and Facebook can complement traditional epidemiologic data and methods such as hospital- or pharmacy-based data, clinical data, focus interviews, and surveys, which are time-consuming [16,17].

In particular, user-generated content and shared health information on social media can serve as an alternative tool for syndromic surveillance [8,18]. Social media health data can accelerate data collection, curation, and analysis. Analyzing user content (especially on Twitter) and tracking information usage patterns such as users' browsing, searching, clicking, or sharing of information regarding health care can reflect the health status, concerns, awareness, and health-related behaviors of the public [18-22]. User-generated content and shared information on social media can also be used for knowledge translation and to increase awareness among policymakers [8].

Investigating the public's communication framing of and approaches to health issues as observed on social media provides insights into the public's thoughts on, perceptions about, and self-disclosures of disease-related symptoms [13]. This can ultimately assist in the development of health intervention strategies and the design of effective campaigns based on public perceptions.

Studies have focused on quantifying the search queries and tracking the volume of health-related information or user-generated content in electronic media using Google Trends, Google Health application programming interface (API), or Google Flu Trends [23,24]. Recent studies in the fields of risk

communication have applied social network analysis to investigate the structure of knowledge and information-sharing networks and multilevel interaction patterns among users [14,25,26]. Studies suggest that network analysis is particularly useful for tracking not only collaborative networks among different stakeholders but also the necessary source distribution during national disasters or emergencies such as earthquakes [14,25,26]. The timely monitoring of risk networks and public conversations on social media can help foster an understanding of stakeholders' perspectives and assist in establishing the policies required for effective risk interventions and resilience, thus, ensuring the effective management of catastrophic events [27-29].

Objectives

We address three research questions (RQs) about Korea's COVID-19 conversations in terms of socially disseminated Twitter messages. First, is there a difference in communication network structure among the four networks generated from four keywords (written in Korean)—Coronavirus, Corona19, Shincheonji, and Daegu—and what are the characteristics of the conversation patterns among users? Second, which news topics and media channels generate the users' interest, and what are their characteristics? Do the most frequently mentioned news topics among the four networks display any differences in media outlet type? Third, what perspectives on news articles are observed from a media organizational point of view? In other words, do news articles with a medically oriented thematic frame have broader spillover effects on the COVID-19 issue in the Twitter context?

Methods

Data Collection

This study evaluates trends in Korea's COVID-19 conversations using Twitter data. Data were collected on February 29, 2020, roughly covering the most recent weeks in the Twitter database. Using the Twitter search API embedded in NodeXL (Social Media Research Foundation) [30], we crawled Twitter users whose recent tweets contained the term "Coronavirus" (코로나바이러스), "corona19" (코로나19), "Shincheonji" (신천지), or "Daegu" (대구) in Korean. These four terms represent the key issues related to COVID-19 in Korea. The tweets were taken from a data set limited to a maximum of around 18,000 tweets. Twitter is one of the largest social media services in the world and provides conversational data, regardless of personal information collection and use consent, that are available free of charge via various automatic methods. Our final sample comprises 43,832 Twitter users and 78,233 relationships, which includes "tweets," "retweets," "replies," and "mentions" from the selected Twitter networks. Once Twitter users post tweets, which can include text, hashtags, images, and URLs, the entire content of the tweets is included when they are retweeted or mentioned by other Twitter users [31].

Social Media Network Analysis and News Channel Classification

This study used three main methodological approaches. It conducted a social media network analysis to determine how COVID-19-related issues circulate on Twitter. The study traced the characteristics of information diffusion regarding COVID-19 on Twitter by generating communication networks composed of all tweets containing any of the search terms ("Coronavirus," "corona19," "Shincheonji," or "Daegu"). A communication network in this study refers to a social network generated by users to communicate to each other. Each node represents a user, and the links between the users refer to a conversation (ie, a retweet, reply to, or mention). Four networks were generated. A network analysis was conducted to identify the multidimensional communication activities between Twitter users and grasp the nature of the information transmission networks composed of entities such as words, hyperlinks, and hashtags. Twitter users are compiled by subgroup using the Clauset–Newman–Moore cluster algorithm and visualized using the Harel–Koren Fast Multiscale layout algorithm [32]. This method has been widely used in communication research on information measurement [33,34]. The study uses NodeXL to calculate various measures quantifying the efficiency of information diffusion and visualizing the structural topography of the network associated with the results of the four search queries as seeds. The network measures include modularity, the number of self-loops, and connected components. Modularity is an indicator of the community structure [35]. The higher the modularity value, the greater the subgroup interconnection. Components in a network analysis refers to a subgraph that represents nodes connected by paths [36]. A self-loop occurs when virtually no one replies to or mentions a tweet.

We then classified the top news items in terms of their media channels. The media outlet of a news article can be regarded as both a form of carrier interface and a means of expression. In Korea, portals are increasingly becoming the primary point of news access. Given Korea's unique news environment, a media organization's presence on a portal offers an interface between its news production and its readership [37]. Furthermore, a professional news agency provides a means of diffusing breaking news quickly and effectively. Media outlets were classified as portal or nonportal news. News stories that were delivered in Korean information portals such as Naver, Daum, and Yahoo! were categorized as portal news. News stories that were not provided through information portals were considered nonportal news. Intercoder reliability scores were calculated based on the coding of 20% (n=8/40) of the sample by a second independent coder using Cohen kappa. There was a strong agreement between the two coders (κ=0.71, P=.03).

Content Analysis of News Frames Used in COVID-19 News Coverage

Besides news channels, we also considered news frames as points of view that guide readers' cognitive direction. Content analysis was conducted to determine the main frames within the news stories by generating content categories that encompass the entire text corpus [38]. News frames were categorized into medical or nonmedical frames. A total of 40 popular news items across the four networks were categorized as "medical vs nonmedical" in terms of their frames. When a news story covered a medical or health issue related to COVID-19, it was classified as a "medical frame."

A coding scheme was developed to further subclassify the nonmedical news frames based on studies of journalists' use of news frames [39,40] and on an inductive review of the news contained in the tweets. The coding scheme was designed to capture how the newspapers promote a specific definition, interpretation, or evaluation of social issues [41]. Nonmedical news frames were classified into five categories: "attribution of responsibility," "human interest," "morality," "entertainment," and "conflict" [39,40]. Textbox 1 presents the definitions of each frame. When more than one frame was used, the dominant frame was selected. Each news item is classified as one frame by one coder (one of the authors). Intercoder reliability scores were computed based on the coding of 20% (n=8/40) of the sample by a second coder using the kappa coefficient. There was a very strong agreement between the two coders (κ=0.837, P<.001).

Textbox 1. Definitions of news frames used in coronavirus disease news coverage.

Conflict

Reflects disagreement between parties, individuals, or groups

Human interest

Emphasizes the (positive) role of individuals and groups affecting the issue

Attribution of responsibility

Suggests that some government agencies, including politicians and public officials, are responsible for the issue

Morality

Contains a moral and ethical message

Medical

Mentions medical and health issues related to the problem

Entertainment

Covers cultural issues such as celebrity, sports, or food

Results

Comparing Communication Network Structures

We addressed RQ1 by comparing the topologies of the four networks, which are shown in Figure 1. The top 10 subgroups are indicated in the graphs. In undirected networks, the size of the radius of the nodes reflects the size of their betweenness centrality values. Due to space constraints, we did not present the ties between nodes.

As Figure 1 shows, the patterns of conversation and information sharing between users are similar across all networks. The results suggest a complex structure composed of multiple layers. Around 20 major subcommunities were activated, and many interactions occurred at an individual level with a small number of isolated users. Table 1 outlines the characteristics of the conversational relationships among users in each network.

The frequency of unique edges was lowest in the Coronavirus network. Unique edges reflect frequency, excluding redundant relationships. In other words, the Coronavirus network has the most redundant relationships, indicating that people continued to talk to each other while exchanging comments several times. Thus, it is highly likely that a "big mouth" existed in the Coronavirus network. On the other hand, the Daegu and Corona19 networks had the lowest frequencies of edges with duplicates. These fewer overlapping relationships suggest that many one-time conversations took place, forming an instant and improvised community.

There are many self-loops in the Shincheonji network, wherein tweets started and ended with the same user in a conversation thread. The Shincheonji network has the highest reciprocated vertex pair and reciprocated edge ratios, which shows that "birds of a feather flock together." When two Twitter users talk to each other, their relationship is regarded as being reciprocated. In sharp contrast, less than one-tenth of tweets in the Coronavirus network were self-loops, revealing that different types of users were paired in comment exchanges. Similarly, the Shincheonji network also had the largest number of isolates, followed by the Daegu network. An isolate has zero connections. This result suggests that the communication patterns in region-oriented networks, Shincheonji and Daegu, differed from the networks that concerned more overall issues regarding COVID-19. The difference can be attributed to geographic variation in information-sharing behaviors on Twitter [42].

Figure 1. Korean coronavirus disease networks on Twitter. Coronavirus (top left), Corona19 (top right), Shincheonji (bottom left), and Daegu (bottom right).

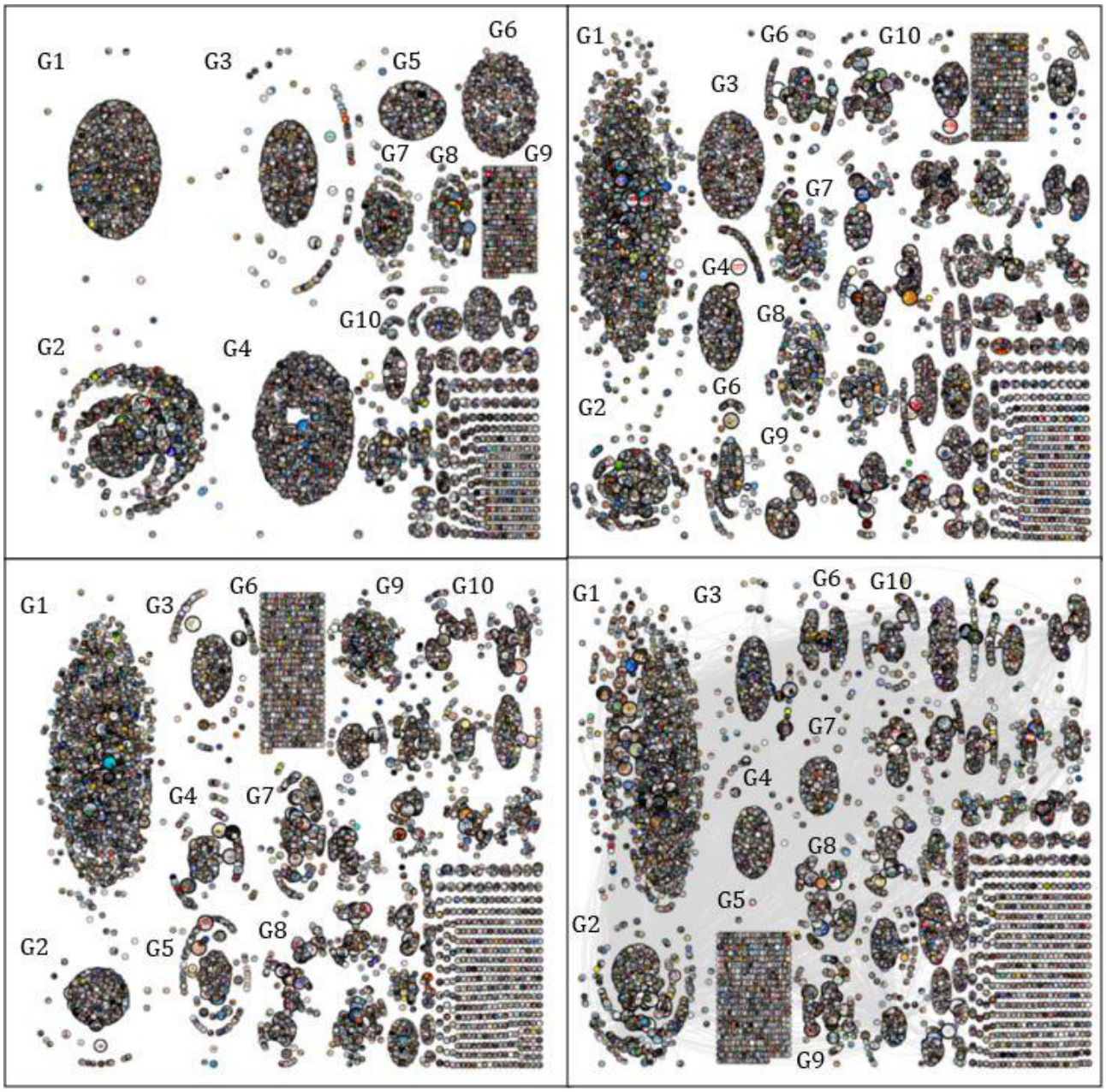


Table 1. Comparing user relationships across coronavirus disease networks.

Network measures	Coronavirus	Corona19	Shincheonji	Daegu
Nodes, n	12,803	11,739	9589	9701
Isolates, n (%)	368 (2.87)	324 (2.76)	471 (4.91)	434 (4.47)
Total edges, n	18,407	19,772	20,327	19,727
Unique edges, n (%)	14,486 (78.70)	17,042 (86.19)	16,879 (83.04)	17,017 (86.26)
Edges with duplicates, n (%)	3921 (21.30)	2730 (13.81)	3448 (16.96)	2710 (13.74)
Self-loops, n (%)	1450 (7.88)	2318 (11.72)	2754 (13.55)	1901 (9.64)
Reciprocated vertex pair ratio	0.00020	0.00042	0.00353	0.00334
Reciprocated edge ratio	0.00040	0.00084	0.00704	0.00666

Table 2 summarizes the overall attributes of each network. When a Twitter network is drawn in concentric form, the step required to connect two users, potentially through intermediate users, reflects the “geodesic” value. The path that connects the furthest

pair is the maximum geodesic distance, or the diameter of the network. The Coronavirus network has the smallest diameter and average values among the four networks. Because people who used the word "Coronavirus" communicated much more frequently with each other than those who used other terms, the spread of information was much faster, and thus, its value is the lowest. By contrast, the Corona19 network has the largest geodesic values.

Next, the modularity value of the Coronavirus network was the highest among the four networks, while the Daegu network had the lowest value. This result suggests that the clusters created within the Coronavirus network may be less cohesive in terms of the subgroups' internal collectivity because the Twitter users in group A tend to be connected with other users in group B. If modularity is low, the clusters are well-defined in terms of the quality of the subgroups generated. Because the Daegu and Shincheonji networks both have lower modularity than the other networks, users classified in the same cluster rarely left their own group to talk to others in a different cluster.

A component analysis reveals that Shincheonji and Daegu network members had the largest numbers of connected components and the maximum edges in a connected component. Coronavirus network members had the largest chat room with the highest value for maximum vertices (ie, users) in a connected component, followed by the users of Corona19.

Table 2. Comparing coronavirus disease network properties on Twitter.

Types	Coronavirus	Corona19	Shincheonji	Daegu
Maximum geodesic distance (diameter)	12	16	14	14
Average geodesic distance	3.865	5.459	4.432	4.471
Modularity	0.674	0.667	0.563	0.530
Connected components, n	584	582	799	828
Maximum vertices in a connected component, n	10,783	10,135	8270	8098
Maximum edges in a connected component, n	16,121	17,916	18,947	18,123

Popular News Topics and Frequently Cited Media Outlets

RQ2 addresses the popularity of news topics and media channels. We extracted the most cited news among the four networks. Five news items appeared twice on the top 10 list of the four Twitter networks. The most popular news concerned suspicions that the *Daily Best*'s online bulletin board deleted all the posts that mentioned Shincheonji. The second most popular news was about the warning given by the Korea Centers for Disease Control that 18 types of beards could be dangerous amid the COVID-19 risk. The third most popular news was the breaking news that the COVID-19 infection rate among ordinary citizens was low in Daegu, unlike for members of Shincheonji. The fourth-ranked news item concerned how messages of support from all over the country, including handwritten letters and fruit, gave Daegu medical staff the strength to continue working. The fifth-ranked concerned news was that college students had created a fact-checking site to prevent the spread of fake news related to COVID-19.

The top 10 news stories on the Twitterverse related to Shincheonji included the term "Shincheonji" in their headline titles. For example, the most popular news was that Shincheonji leader Lee Man-hee had received recognition for his service to the country from ex-President Park Geun-hye and that he was set to be buried at the National Cemetery. Thus, the top news stories shared on Twitter featured eye-catching headlines, the use of dramatic expressions, and emotional narratives.

It is noteworthy that international news, rather than domestic news, was chosen as the top news item. *Foreign Policy* (FP), a US diplomatic magazine, diagnosed the cause of Korea's COVID-19 problem [43]. The article was titled "Cults and Conservatives Spread Coronavirus in South Korea: Seoul Seemed to Have the Virus under Control, but Religion and Politics Have Derailed Plans." Additionally, a petition posted to the presidential "Blue House" was also a top news item [44]. The petition called for the dismissal of Prosecutor General Yoon Seok-Yeol for failure of leadership because he did not deal promptly with the Shincheonji-related situation despite the deep concerns of the Korean public.

Finally, we investigated the most cited news channels among the four networks. Korea's portals and *Yonhap News* (Korea's largest news agency, comparable to the US Associated Press) were ranked highly in frequently linked Twitter groups (n=26 and n=8, respectively). The preferred channel of popular news in the Coronavirus Twitter network was *Yonhap News*. Among the portals, *Daum*, the second largest portal in Korea, was overwhelmingly more popular than *Naver*, the largest portal in Korea, (n=22 and n=4, respectively). The rest include online-focused newspapers (n=2), social media postings (n=2, with one tweet pointing to the FP article and the other citing the famous "Real-time in Daegu" Facebook page), a personal blog (n=1), and a petition site (n=1). The dominance of the portals and *Yonap News* is attributable to their business model of delivering news faster than other media outlets do. Surprisingly, no traditional media outlet appears among the top news channels.

The study computed a 4 (Corona19 vs Coronavirus vs Daegu vs Shincheonji) x 2 (portal vs nonportals) chi-square comparing the frequency of portal vs nonportal news site use between network types. The difference is found to be significant (χ^2_3=12.747, *P*=.005). The results indicated that the Corona 19 network cited more portal news (n=8/10, 80%) than did nonportal news (n=2, 20%), whereas the Coronavirus network used more nonportal sources (n=8/10, 80%) than portal news

(n=2, 20%), *P*=.007. The Coronavirus network also used more nonportal sources than portal sources, while the Daegu network (*P*=.002) and Shincheonji network (*P*=.02) included more portal news (n=9/10, 90% for Daegu; n=7/10, 70% for Shincheonji) than nonportal sources (n=1, 10% for Daegu; n=3, 30% for Shincheonji).

We also determined the origin of the news items reported in the portals. The four *Naver* news articles were based on *SEN*, the *Sports Donga* newspaper (both online-focused newspapers), *Newsis* (a news agency), and *Maeil Broadcasting Network* (a cable news channel, comparable to CNN). Some 22 *Daum* news articles came from the *News1* (news agency; n=3); *Newsis* (n=2); *OhMyNews*, an online-focused newspaper (n=1); *Nocut* News, an online-focused newspaper (n=4); the *Seoul Daily* newspaper (n=2); the *World Daily* newspaper (n=1); *Yonhap News* (n=4); the *JoongAng Daily* newspaper (n=2); the *Hankyoreh Daily* newspaper (n=2); and the *Hankook Daily* newspaper (n=1). The so-called "legacy" media are eager to provide their news via Korea's portals because that enables their breaking and exclusive news articles to attract greater attention from the portals' readership.

News Frames and Popularity of News

This study addresses RQ3 by analyzing the news frames of media organizations that were circulated in tweets. The results of content analysis show that 17.5% (n=7/40) of the articles mentioned medical or health problems. The medical news items include discussions of the characteristics of COVID-19, warnings about the potential for beards to be infected with COVID-19, the effects of health conditions on mortality, the status of COVID-19 tests in Italy, and the difference in infection rates between Shincheonji members and ordinary citizens.

The results of an independent two-tailed *t*-test suggest that tweets containing medically framed news articles (mean 7.571, SD 1.988) are found to be more popular than tweets that included news articles adopting nonmedical frames (mean 5.060, SD 2.904, *P*=.03, 95% CI 0.169-4.852).

This study ranked the most popular news articles from 1 to 10. This study then calculated reverse scores to measure the spillover effects of the articles. For example, the top-ranked news item were given 10 points, and the 10th-ranked item was given 1 point.

Lastly, the study investigated the news frames included in the tweets produced across the four networks and compared the association between the network typology and the frames. The findings suggest that the "attribution of responsibility" frame was the most frequently used, followed by "human interest." Both "morality" and "entertainment" were cited 6 times. "Conflict" was the least used frame. This study conducted a 4 (network types) x 6 (news frames) chi-square analysis to examine the association between network type and the news frames used in the tweets. As shown in Table 3, the results indicate that network type was not significantly associated with the news frames (*P*=.89, 95% CI 0.004-0.006), suggesting that the six frames were used similarly across the four networks.

Table 3. Chi-square results for the news frames across coronavirus disease networks.

Network type	Total (N=40), n (%)	Conflict (n=4), n (%)	Entertainment (n=6), n (%)	Human interest (n=7), n (%)	Medical (n=7), n (%)	Morality (n=6), n (%)	Attribution of responsibility (n=11), n (%)	Chi-square (*df*)
All networks	N/A[a]	N/A	N/A	N/A	N/A	N/A	N/A	8.727 (15)
Corona19	10 (25)	1 (10)[b]	2 (20)	2 (20)	3 (30)	1 (10)	1 (10)	N/A
Coronavirus	10 (25)	0 (0)	1 (10)	2 (20)	2 (20)	2 (20)	3 (30)	N/A
Daegu	10 (25)	0 (0)	2 (20)	2 (20)	1 (10)	1 (10)	4 (40)	N/A
Shincheonji	10 (25)	2 (20)	1 (10)	1 (10)	1 (10)	2 (20)	3 (30)	N/A

[a]Not applicable.

[b]In the calculation of the n (%) values across each row, the row total is taken as the N value.

Discussion

Principal Findings

By March 1, 2020, Korea had become one of the most SARS-CoV-2-infected countries in the world. The greater Daegu metropolitan area had Korea's highest COVID-19 infection rate per household as well as the highest absolute rate [45]. This study has found that the spread of information was faster in the Coronavirus network than in the other networks because Twitter users in cluster A created within that network tended to be interconnected with others in cluster B. These findings have implications for risk communication and health campaigns. When government authorities and experts share and comment on real time information about ongoing infectious disease threats, social media analytics can help them to choose what keywords and hashtags are more appropriate to use. In fact, central and regional governments in Korea have been criticized for causing social confusion by failing to effectively communicate proper information on how to deal with the symptoms of infectious diseases. Our findings can be used to enhance government fact-checking services and risk communication. For example, the frequently shared topics and information sources regarding the Shincheonji-related Twitter networks could help both local and central governments to review people's concerns and opinions regarding the case and to guide information strategies in epidemic situations, which often demand prompt decision making for risk management.

Our main research platform was Twitter, but the analysis has also considered intermedia journalism that goes beyond Twitter. People use various news channels to share information, even

when communicating via social media. This study found that portals were the preferred news sources on Twitter. As shown in Endo's [46] study on Japan's 3.11 earthquake, various media interact with each other during national disasters, creating a social information environment that can be both positive and negative. It would be ideal if intermedia journalism created an information immunization system during epidemics. This study is important in that it examines a key aspect of intermedia journalism; although, the full dynamics of intermedia journalism during Korea's COVID-19 crisis has yet to be fully investigated.

Limitations

Although this study proposed and demonstrated a useful infodemiology framework by performing social network analytics to explore information diffusion related to the COVID-19 pandemic via Twitter in Korea, it is not without limitations. These limitations are inherent in Twitter's user population. The literature suggests that only 15% of online adults are regular Twitter users [47]. Moreover, the largest group among Twitter users is composed of those 18-29 years of age. In addition, only a small number of Twitter users are active in producing tweets and leveraging the discourse [47-49]. There are more very passive (<50 tweets per year) and very active users (>1000 tweets per year) on Twitter than moderate users (50-1000 tweets a year) [46]. Thus, the study's results may reflect social media users' views and behaviors during the pandemic rather than the full population's aggregate opinion. In addition, biases in information-sharing behaviors can exist, as some users may have produced more content than others. It was also possible that whether they are residents or visitors of Daegu City may have influenced what news they shared [50]. Furthermore, we qualitatively examined the study's data sets to detect abnormal activities such as data contamination from social bots; however, running Twitter bot detection software would have enabled us to more systematically remove any potential biases from the data set. Finally, we applied multi-coder methods, but two coders may not be sufficient to ensure reliability. Future studies could apply more coders to guarantee the reliability of the analysis results.

Conclusions

People experiencing social disasters such as epidemics of infectious diseases are unfamiliar with their situation and find it difficult to predict what will happen next. Therefore, risk communication that delivers accurate and appropriate information is important. We found that most of the popular news on Korea's Twitterverse had nonmedical frames. Nevertheless, it must be noted that the spillover effect of the news articles that delivered medical information about COVID-19 was greater than that of news with nonmedical frames. For instance, many news items reported that the initial response of government agencies was responsible for the spread of COVID-19. Many news items highlighted the positive role of individuals and groups, directing readers' attention to the epidemic crisis. Ethical issues such as deviant behavior among the population and an entertainment frame highlighting celebrity donations also emerged often. Relatively few articles reflected discrepancies in positions or opinions among individuals or groups.

Augmented intelligence systems in the medical sector have been widely cited as an important approach to helping detect and clinically diagnose diseases [51]. However, this study is not intended to provide public health care data to artificial intelligence systems, such as HealthMap or BlueDot, to assist in the prediction of infectious disease occurrence. Rather, this study emphasizes the need and opportunity for strategic communication through social media. Social media network analytics cannot replace the work of public health officials; however, collecting public conversations and media news that propagates rapidly and detecting its structure can assist public health professionals in their complex and fast-paced decision-making processes.

Acknowledgments

HWP would like to thank Dr Marc Smith and the Social Media Foundation for offering a valuable Twitter data set via NodeXL. SP appreciates the financial support of John Carroll University.

Conflicts of Interest

None declared.

References

1. Wong T. BBC.: www.bbc.com; 2020 Mar 14. Shincheonji and coronavirus: the mysterious 'cult' church blamed for S Korea's outbreak URL: https://www.bbc.com/news/av/world-asia-51851250/shincheonji-and-coronavirus-the-mysterious-cult-church-blamed-for-s-korea-s-outbreak
2. Rashid R. The New York Times. 2020 Mar 09. Being called a cult is one thing, being blamed for an epidemic is quite another URL: https://www.nytimes.com/2020/03/09/opinion/coronavirus-south-korea-church.html
3. Hern A. The Guardian. 2020 Mar 04. Fake coronavirus tweets spread as other sites take harder stance URL: https://www.theguardian.com/world/2020/mar/04/fake-coronavirus-tweets-spread-as-other-sites-take-harder-stance
4. Evon D. Snopes. 2020 Feb 14. Does video show guns, violence in aftermath of coronavirus outbreak in China? URL: https://www.snopes.com/fact-check/violent-aftermath-of-coronavirus/
5. Park HW. YouTubers' networking activities during the 2016 South Korea earthquake. Qual Quant 2017 Mar 27;52(3):1057-1068. [doi: 10.1007/s11135-017-0503-x]

6. Shearer E, Gottfried J. Pew Research Center. 2017 Sep 07. News use across social media platforms URL: https://www.journalism.org/2017/09/07/news-use-across-social-media-platforms-2017/
7. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11. [doi: 10.2196/jmir.1157] [Medline: 19329408]
8. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):S154-S158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]
9. Alicino C, Bragazzi NL, Faccio V, Amicizia D, Panatto D, Gasparini R, et al. Assessing Ebola-related web search behaviour: insights and implications from an analytical study of Google Trends-based query volumes. Infect Dis Poverty 2015 Dec 10;4:54. [doi: 10.1186/s40249-015-0090-9] [Medline: 26654247]
10. Bragazzi NL, Watad A, Brigo F, Adawi M, Amital H, Shoenfeld Y. Public health awareness of autoimmune diseases after the death of a celebrity. Clin Rheumatol 2017 Aug;36(8):1911-1917. [doi: 10.1007/s10067-016-3513-5] [Medline: 28000011]
11. Ayers JW, Althouse BM, Allem J, Ford DE, Ribisl KM, Cohen JE. A novel evaluation of World No Tobacco day in Latin America. J Med Internet Res 2012 May 28;14(3):e77. [doi: 10.2196/jmir.2148] [Medline: 22634568]
12. Lorence D, Park H. Group disparities and health information: a study of online access for the underserved. Health Informatics J 2008 Mar;14(1):29-38. [doi: 10.1177/1460458207086332] [Medline: 18258673]
13. Park S, Hoffner CA. Tweeting about mental health to honor Carrie Fisher: how #InHonorOfCarrie reinforced the social influence of celebrity advocacy. Computers in Human Behavior 2020 Mar:106353. [doi: 10.1016/j.chb.2020.106353]
14. Cho SE, Jung K, Park HW. Social media use during Japan's 2011 earthquake: how Twitter transforms the locus of crisis communication. Media International Australia 2013 Nov;149(1):28-40. [doi: 10.1177/1329878x1314900105]
15. Zhao Y, Zhang J. Consumer health information seeking in social media: a literature review. Health Info Libr J 2017 Dec;34(4):268-283. [doi: 10.1111/hir.12192] [Medline: 29045011]
16. Aslam AA, Tsou M, Spitzberg BH, An L, Gawron JM, Gupta DK, et al. The reliability of tweets as a supplementary method of seasonal influenza surveillance. J Med Internet Res 2014 Nov 14;16(11):e250. [doi: 10.2196/jmir.3532] [Medline: 25406040]
17. Yang YT, Horneffer M, DiLisio N. Mining social media and web searches for disease detection. J Public Health Res 2013 Apr 28;2(1):17-21. [doi: 10.4081/jphr.2013.e4] [Medline: 25170475]
18. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. In: AMIA Annu Symp Proc. 2006 Presented at: AMIA Annual Symposium; November 11, 2006 - November 15, 2006; Washington, D.C p. 244-248 URL: http://europepmc.org/abstract/MED/17238340
19. Achrekar H, Gandhe A, Lazarus R, Yu S, Liu B. Twitter improves seasonal influenza prediction. Healthinf 2012 Feb:61-70. [doi: 10.5220/0003780600610070]
20. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. PLoS One 2011 May 04;6(5):e19467. [doi: 10.1371/journal.pone.0019467] [Medline: 21573238]
21. Davis MA, Zheng K, Liu Y, Levy H. Public response to Obamacare on Twitter. J Med Internet Res 2017 May 26;19(5):e167. [doi: 10.2196/jmir.6946] [Medline: 28550002]
22. Hart M, Stetten NE, Islam S, Pizarro K. Twitter and public health (part 1): how individual public health professionals use Twitter for professional development. JMIR Public Health Surveill 2017 Sep 20;3(3):e60. [doi: 10.2196/publichealth.6795] [Medline: 28931499]
23. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
24. Arora VS, McKee M, Stuckler D. Google Trends: opportunities and limitations in health and health policy research. Health Policy 2019 Mar;123(3):338-341. [doi: 10.1016/j.healthpol.2019.01.001] [Medline: 30660346]
25. Jung K, Park HW. Tracing interorganizational information networks during emergency response period: a webometric approach to the 2012 Gumi chemical spill in South Korea. Government Information Quarterly 2016 Jan;33(1):133-141. [doi: 10.1016/j.giq.2015.09.010]
26. Park S. We love or hate when celebrities speak up about climate change: receptivity to celebrity involvement in environmental campaigns. J Contemp East Asia 2019;18(1):175-188. [doi: 10.17477/JCEA.2019.18.1.175]
27. Jung K, Song M, Park HW. Filling the gap between bureaucratic and adaptive approaches to crisis management: lessons from the Sewol Ferry sinking in South Korea. Qual Quant 2017 Jan 18;52(1):277-294. [doi: 10.1007/s11135-017-0467-x]
28. Song MS, Jung KJ, Kim JY, Park HW. Risk communication on social media during the Sewol Ferry disaster. J Contemp East Asia 2019;18(1):173-200.
29. Jung K, Park HW. Citizens' social media use and homeland security information policy: some evidences from Twitter users during the 2013 North Korea nuclear test. Government Information Quarterly 2014 Oct;31(4):563-573. [doi: 10.1016/j.giq.2014.06.003]
30. Smith MA. Catalyzing social media scholarship with open tools and data. J Contemp East Asia 2015 Oct 31;14(2):87-96. [doi: 10.17477/jcea.2015.14.2.087]

31. Chong M. Sentiment analysis and topic extraction of the twitter network of #prayforparis. In: Proc Assoc Info Sci Tech. 2016 Dec 27 Presented at: Proceedings of the Association for Information Science and Technology ; 53(1)?4; 2016; Copenhagen p. 1-4. [doi: 10.1002/pra2.2016.14505301133]
32. Clauset A, Newman M, Moore C. Finding community structure in very large networks. Phys Rev E Stat Nonlin Soft Matter Phys 2004 Dec;70(6 Pt 2):066111. [doi: 10.1103/PhysRevE.70.066111] [Medline: 15697438]
33. Park S, Chung D, Park HW. Analytical framework for evaluating digital diplomacy using network analysis and topic modeling: comparing South Korea and Japan. Information Processing & Management 2019 Jul;56(4):1468-1483. [doi: 10.1016/j.ipm.2018.10.021]
34. Park HW, Yoon J. Structural characteristics of institutional collaboration in North Korea analyzed through domestic publications. Scientometrics 2019 Mar 9;119(2):771-787. [doi: 10.1007/s11192-019-03056-5]
35. Newman MEJ. Modularity and community structure in networks. In: Proc Natl Acad Sci U S A. 2006 Jun 06 Presented at: Proc Natl Acad Sci U S A. ;103(23); 2006; Irvine p. 8577-8582 URL: http://www.pnas.org/cgi/pmidlookup?view=long&pmid=16723398 [doi: 10.1073/pnas.0601602103]
36. Wasserman S, Faust K. Social Network Analysis: Methods and applications (Vol 8). Cambridge: Cambridge University Press; 1994.
37. Dwyer T. Special issue: Media manipulation, fake news, and misinformation in the Asia-Pacific region. J Contemp East Asia 2019;18(2):9-15.
38. Diallo SY, Gore RJ, Padilla JJ, Lynch CJ. An overview of modeling and simulation using content analysis. Scientometrics 2015 Mar 26;103(3):977-1002. [doi: 10.1007/s11192-015-1578-6]
39. Gonzálo BS, García RM, Chico TLM. Economic crisis as spectacle in Spain: infotainment in quality press coverage of the 2012 financial sector rescue. Communication & Society 2015 Oct 15;28(4):1-16. [doi: 10.15581/003.28.4.1-16]
40. Valkenburg PM, Semetko HA, De Vreese CH. The effects of news frames on readers' thoughts and recall. Communication Research 2016 Jun 30;26(5):550-569. [doi: 10.1177/009365099026005002]
41. Entman RM. Framing: toward clarification of a fractured paradigm. Journal of Communication 1993 Dec;43(4):51-58. [doi: 10.1111/j.1460-2466.1993.tb01304.x]
42. Gore RJ, Diallo S, Padilla J. You are what you tweet: connecting the geographic variation in America's obesity rate to Twitter content. PLoS One 2015;10(9):e0133505. [doi: 10.1371/journal.pone.0133505] [Medline: 26332588]
43. Foreign Policy. 2020 Feb 27. Cults and conservatives spread coronavirus in South Korea URL: https://foreignpolicy.com/2020/02/27/Coronavirus-south-korea-cults-conservatives-china/ [accessed 2020-02-28]
44. President.go.kr. 2020 Feb 28. "Shincheonji seizure and search must be cautious," the prosecutor-general Yoon Seok-yeol demands URL: https://www1.president.go.kr/petitions/585895
45. Hermesauto T. The Straits Times. 2020 Mar 01. Coronavirus: South Korea wages 'all-out responses' to virus, with 586 new cases URL: https://www.straitstimes.com/asia/east-asia/coronavirus-south-korea-reports-210-new-cases-raising-total-to-3736
46. Endo K. Social journalism in the inter-media society: results from the social survey on the great East Japan earthquake and the Fukushima Nuclear Power Plant disaster. J Contemp East Asia 2013 Oct 31;12(2):5-17. [doi: 10.17477/jcea.2013.12.2.005]
47. Docherty M, Smith R. The case for structuring the discussion of scientific papers. BMJ 1999 May 08;318(7193):1224-1225 [FREE Full text] [doi: 10.1136/bmj.318.7193.1224] [Medline: 10231230]
48. Park SJ, Park JY, Lim YS, Park HW. Expanding the presidential debate by tweeting: the 2012 presidential election debate in South Korea. Telematics and Informatics 2016 May;33(2):557-569. [doi: 10.1016/j.tele.2015.08.004]
49. Chong M, Kim HJ. Social roles and structural signatures of top influentials in the #prayforparis Twitter network. Qual Quant 2019 Nov 15;54(1):315-333. [doi: 10.1007/s11135-019-00952-z]
50. Padilla JJ, Kavak H, Lynch CJ, Gore RJ, Diallo SY. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. PLoS One 2018 Jun 14;13(6):e0198857. [doi: 10.1371/journal.pone.0198857] [Medline: 29902270]
51. Long JB, Ehrenfeld JM. The role of augmented intelligence (AI) in detecting and preventing the spread of novel coronavirus. J Med Syst 2020 Feb 04;44(3):59 [FREE Full text] [doi: 10.1007/s10916-020-1536-6] [Medline: 32020374]

Abbreviations

API: application programming interface
COVID-19: coronavirus disease
FP: Foreign Policy
RQ: research question
SARS-CoV-2: severe acute respiratory coronavirus 2
SNS: social networking services

Edited by G Eysenbach; submitted 26.03.20; peer-reviewed by JP Allem, R Gore; comments to author 05.04.20; revised version received 16.04.20; accepted 22.04.20; published 05.05.20.

Please cite as:
Park HW, Park S, Chong M
Conversations and Medical News Frames on Twitter: Infodemiological Study on COVID-19 in South Korea
J Med Internet Res 2020;22(5):e18897
URL: http://www.jmir.org/2020/5/e18897/
doi:10.2196/18897
PMID:32325426

Original Paper

Containing COVID-19 Among 627,386 Persons in Contact With the Diamond Princess Cruise Ship Passengers Who Disembarked in Taiwan: Big Data Analytics

Chi-Mai Chen[1], MD, MSc; Hong-Wei Jyan[2], MSc; Shih-Chieh Chien[2], PhD; Hsiao-Hsuan Jen[3], PhD; Chen-Yang Hsu[3], PhD, MD; Po-Chang Lee[4], MD, MSc; Chun-Fu Lee[4], MSc; Yi-Ting Yang[5], MSN; Meng-Yu Chen[5], MD, MSc; Li-Sheng Chen[6], PhD; Hsiu-Hsi Chen[3], PhD; Chang-Chuan Chan[7], ScD

[1]Executive Yuan, Taipei, Taiwan

[2]Department of Cyber Security, Executive Yuan, Taipei, Taiwan

[3]Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan

[4]National Health Insurance Administration, Ministry of Health and Welfare, Taipei, Taiwan

[5]Centers for Disease Control, Ministry of Health and Welfare, Taipei, Taiwan

[6]School of Oral Hygiene, College of Oral Medicine, Taipei Medical University, Taipei, Taiwan

[7]Institute of Environmental and Occupational Health Sciences, College of Public Health, National Taiwan University, Taipei, Taiwan

Corresponding Author:
Chang-Chuan Chan, ScD
Institute of Environmental and Occupational Health Sciences
College of Public Health
National Taiwan University
No.17, Xuzhou Road
Taipei, 100
Taiwan
Phone: 886 233668082
Email: ccchan@ntu.edu.tw

Abstract

Background: Low infection and case-fatality rates have been thus far observed in Taiwan. One of the reasons for this major success is better use of big data analytics in efficient contact tracing and management and surveillance of those who require quarantine and isolation.

Objective: We present here a unique application of big data analytics among Taiwanese people who had contact with more than 3000 passengers that disembarked at Keelung harbor in Taiwan for a 1-day tour on January 31, 2020, 5 days before the outbreak of coronavirus disease (COVID-19) on the Diamond Princess cruise ship on February 5, 2020, after an index case was identified on January 20, 2020.

Methods: The smart contact tracing–based mobile sensor data, cross-validated by other big sensor surveillance data, were analyzed by the mobile geopositioning method and rapid analysis to identify 627,386 potential contact-persons. Information on self-monitoring and self-quarantine was provided via SMS, and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) tests were offered for symptomatic contacts. National Health Insurance claims big data were linked, to follow-up on the outcome related to COVID-19 among those who were hospitalized due to pneumonia and advised to undergo screening for SARS-CoV-2.

Results: As of February 29, a total of 67 contacts who were tested by reverse transcription–polymerase chain reaction were all negative and no confirmed COVID-19 cases were found. Less cases of respiratory syndrome and pneumonia were found after the follow-up of the contact population compared with the general population until March 10, 2020.

Conclusions: Big data analytics with smart contact tracing, automated alert messaging for self-restriction, and follow-up of the outcome related to COVID-19 using health insurance data could curtail the resources required for conventional epidemiological contact tracing.

(J Med Internet Res 2020;22(5):e19540) doi:10.2196/19540

KEYWORDS

COVID-19; mobile geopositioning; contact tracing; proximity tracing; digital contact tracking; big data; public health; precision public health; surveillance; virus

Introduction

Taiwan has been acclaimed for a relatively low number of coronavirus disease (COVID-19) confirmed cases and case-fatality rates by its timely and fast response to the COVID-19 pandemic [1]. Taiwan activated the Central Epidemic Command Center (CECC) for the COVID-19 outbreak after the first case was confirmed on January 21, 2020, in Taiwan, and this center responsible for executing control policies including border control, surveillance, quarantine, and resource allocation to prevent the spread of COVID-19 in communities [2]. The crucial factor that renders these control measures successful is comprehensive, precise, and timely contact tracing to identify and manage the potential secondary cases and to interrupt further onward transmission. However, the conventional epidemiological contact tracing, which relies on personal interviews is labor-intensive and time-consuming, may not be feasible when dealing with a pandemic with rapid propagation such as COVID-19. To enable more efficient and effective contact investigation, several digital databases such as electronic health records, phone-based GPS, card transaction, records, and closed-circuit television have been applied in South Korea [3]. Other contributory factors that are effective and efficient to contain transmission are quarantine, isolation, and surveillance of disease progression of COVID-19 after contact tracing. To achieve these two aims, a systematic and efficient big data method, using digital technology, sensor data, and claimed health insurance data, may strengthen the conventional contact tracing and disease surveillance and inform the following control measures or mitigation plan. The scientific society in Taiwan thus called for an innovative and integrated approach by making use of current digital technologies and big data on sensor and claimed health insurance to reach the aim of precision prevention for outbreak and surveillance of disease outcome among these contacts.

Here, we present a unique example of retrospectively investigating a substantial proportion of people who had contact with the passengers of the Diamond Princess cruise ship docked at Keelung, Taiwan, on January 31, 2020; the mitigation plan using self-isolation and self-monitoring syndromes of COVID-19; and the disease surveillance of those contacts using efficient big data analysis to link those contacts with the National Health Insurance Claims data in order to ascertain presence of COVID-19 and related respiratory syndrome. Considering the incubation period of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the CECC decided to implement additional precautionary measures to further reduce the risk of importation of COVID-19 to Taiwan, as the time the passengers lingered around northern Taiwan was 5 days before the outbreak of COVID-19 was reported from the Diamond Princess cruise ship at Yokohama on February 5, 2020.

Methods

Overview

The first patient who had returned from Wuhan to Taiwan was confirmed to have COVID-19 on January 21, 2020, in Taiwan. After that, the CECC was activated by the Taiwanese government to control the possible outbreak of COVID-19 [2]. The CECC is responsible for making and executing policies including surveillance, border control, quarantine, and resource allocation and announced the highest level of alarm to prevent the outbreak of COVID-19 in Taiwan.

Although less than 20 cases were reported during the first COVID-19 epidemic period (from January 21 to February 9), the government has paid close attention to all possible leaks responsible for the transmission of COVID-19. The implementation strategies included border control, quarantine, and isolation. The first entry restriction on foreigners from pandemic areas such as China in response to COVID-19 was initiated on January 28, 2020. The government also kept watch on the cruise ships coming to Taiwan, which included the Diamond Princess cruise ship docked at Keelung harbor in Taiwan, on January 31, 2020. Considering the coronavirus incubation period, the CECC decided to implement additional precautionary measures to further reduce the risk of importation of COVID-19 to Taiwan when the outbreak of COVID-19 was reported from the Diamond Princess cruise ship at the Yokohama since it docked at the harbor on February 5, 2020. This unexpected event created a temporary public panic about community spread [4]. Comprehensive contact tracing and a mitigation plan could be some of the strategies to minimize the spread of COVID-19.

Big Data Analytics for Containing the Spread of SARS-COV-2

After knowing about the outbreak of the Diamond Princess cruise ship on February 5, 2020, the CECC immediately formed a task force to involve the preliminary investigation on February 6, 2020. Contact tracing for those possibly contacted by already infected passengers was recommended. The design and process of contact investigation and management were elaborated as follows.

Big Sensor Data to Explore Passengers' Routes

As the cruise ship passengers had a 1-day excursion on January 31, 2020, when the Diamond Princess cruise ship was docked at Keelung harbor, the team designed possible solutions for tracing their routes through their itinerary in Taiwan. As it was impossible to conduct retrospective individual interviews for each passenger, the methods used to overcome the barrier of determining the location and itinerary of the contact were classified into four main categories: GPS in the shuttle bus, credit card transaction log, closed-circuit television (CCTV), and mobile position data.

Among the four categories, the mobile geopositioning method was the mainstay for identifying passengers' routes by mobile position data for COVID-19 contact investigations and was able to provide more accurate information on the location and time of exposure. This method can overcome the shortcoming of incomplete information obtained from the GPS in the shuttle bus, card transactions, and CCTV, as these three methods were only representative of some passengers. These three methods were used for cross-validation of the routes estimated by the mobile sensors of the contacted persons in the light of mobile position data from the passengers.

The mobile position data from more than 3000 passengers on January 31, 2020, were obtained from five local mobile phone companies. The mobile position data are collected at mobile positioning measurements up to 150 meters from the true mobile location as the accuracy of geolocation for identifying possible contact persons. The mobile position method might not be as exact as GPS but the latter may infringe on individual confidentiality. The contact locations were ascertained on the basis of the roaming signals with time of exposure over 30 minutes from multimobile base stations between 5 AM and 8 PM that were recognized as the major tracking routes. Based on the mobile signal registered to the base stations of five domestic telecom operators, the first challenge was to identify the 3000 passengers out of all tourists in Keelung area. According to the record, the cruise was moored at the harbor from 6 AM to 6 PM. We then checked the data between 1 hour before and 2 hours after the cruise docked at Keelung harbor. This confirmed the exact mobile phone numbers of people who traveled with the cruise.

After collecting those phone numbers, the team depicted rough locations of those phones. With the assistance of the local government, we found that about 34% of passengers took shuttle buses for local tours, 5.2% took taxies, the others biked or walked around at harbor or nearby area. More than 24 buses and 50 taxies had been interviewed and recorded. The estimated routes of passengers were further validated by the itineraries provided by the travel agency. The team then checked the detail tour information for each route, interviewed the taxi drivers in harbor area for destination, and integrated all information to confirm more precisely the location where passengers stayed.

The most important part of this stage was to identify the possible position where passengers were. This also showed how to utilize big data analysis with a mixture of different data sources.

Mobile Sensor Data for Identifying the Possible Contacts

At the second stage, we resorted to the mobile position information of passengers above to identify the sensors of mobiles from the possible contact persons. Citizens who carried their mobile phone and stayed within 500 meters of the marked locations over 5 minutes were classified as people who possibly contacted the passengers of the Diamond Princess cruise ship on January 31, 2020.

Sending a Message for Self-quarantine and Self-monitoring to Potential Contacts

On February 7, 2020, the CECC sent an alert notice using SMS through the Public Warning System to remind the contact persons of starting the mitigation plan. The potential contact persons were advised to be quarantine at home, so that they did not engage in public gatherings, to avoid further contact. They were also notified to self-monitor COVID-19–compatible symptoms (fever, cough, and shortness of breath) and seek medical attention when symptoms developed.

Management of Potential Contacts with Symptoms

On February 9, the CECC sent a notice to all health care providers mentioning this event and the guidance for management of potential contacts. Health care professionals were advised to perform SARS-CoV-2 testing for symptomatic contacts. After testing, symptomatic contacts may have been hospitalized as indicated or returned home for self-isolation. Health care professionals were also advised to proactively contact public health authorities to initiate active follow-up of the contacts.

COVID-19 Surveillance for Contact Population Using National Health Insurance Claims Data

In order to capture those in the contact population who sought medical attention but did not report to public health authorities, we used the National Health Insurance Claims data to track the health status of all subjects with potential contact. Those who were hospitalized due to pneumonia were identified. For those who remained hospitalized but had not been tested for SARS-CoV-2, the health care providers were informed of the potential exposure of the patient and screening for SARS-CoV-2 was suggested.

Big Data Analysis for Hospitalization of Patients With Pneumonia Without Reverse Transcription–Polymerase Chain Reaction Test for COVID-19 Through National Health Insurance Claimed Data

As few asymptomatic patients that may have a long duration of COVID-19 development and were very difficult to be identified by the reverse transcription–polymerase chain reaction (RT-PCR) test, it is also very interesting to compare the difference in the rate of respiratory syndrome and pneumonia between the contact population (n=627,386 residents) and the general population in Taiwan (n=23,877,447 residents). Among these subjects, information on respiratory syndrome or pneumonia cases was ascertained by linkage with the big National Health Insurance claim database from January 31, 2020, to March 10, 2020. During this period, subjects with at least one outpatient visit with ICD-10 (The International Statistical Classification of Diseases and Related Health Problems, 10th Revision) codes ("J00" to "J11") were identified as having respiratory syndrome. The subjects who had pneumonia were identified by ICD-10 codes ("J12-" to "J18").

Ethical Considerations

Under the Taiwan Infectious Disease Control Act that was mandated in the year 2007, four years after the outbreak of SARS, authorization or consent to the retrieval of individual information pertaining to containing the outbreak of disease under the auspice of the government can be waived in the face of emerging infectious diseases, such as SARS-COV-2.

Statistical Analysis

In order to evaluate whether there was a significant increase in the rate of respiratory syndrome as well as the rate of pneumonia after sending the alert message to those in the route of passenger-visited areas, we compared these rates between the contact population and the general population. The age-standardized rates between the two groups were calculated. The Breslow and Day method [5] was used to calculate 95% confidence intervals.

Results

Analysis and Decision of Digital Contact Tracing

Multiple means were taken for contact tracing by the CECC. These included travelling itinerary arranged by the agency, GPS in shuttle buses, credit card transaction logs, CCTV, vehicle license plate recognition system, and mobile positioning data (Table 1). The pros and cons of these means are delineated below:

1. Traveling itinerary: The traveling itinerary proposed by the travel agency provided information on the schedule and places of visit. However, this method can only trace some passengers. In addition, the specific time for visiting a place from the itinerary is sometimes not reliable.
2. GPS in shuttle buses: GPS route records in the shuttle buses were considered. However, shuttle buses were used only by some passengers. Others could travel by other modes of transportation.
3. Credit card transaction log: The advantage of using credit card transaction logs is the specificity of the individual, time, and space. The difficulty is the accessibility of these data. Even if it is feasible, this data set cannot trace those who did not use credit cards.
4. CCTV and vehicle license plate recognition system: To trace the routes of shuttle buses or private transportations, CCTVs can target a specific vehicle or passenger. However, the coverage of CCTV is not 100%. Besides, the large number of passengers makes the tracing with CCTV and license plate recognition impossible.
5. Highway electronic toll collection system: All vehicles passing the national freeways were checked with the electronic toll collection system in Taiwan. It can trace a specific person or vehicle, but is not feasible when the number of passengers is large.
6. Mobile positioning data: Passengers travelling with a mobile phone can be traced with mobile positioning service for a specific time and space. In addition, the same information can be applied to delivering a warning message to citizens who were potentially in contact with the passengers. However, this method may miss those who did not carry a mobile phone, but this is a very rare scenario in Taiwan.

Finally, considering the specificity, feasibility, and largest coverage of passengers, the CECC made a final decision to use mobile positioning data for contact tracing and delivery of alert messages. It should be noted that other methods, except the credit card transaction due to privacy concern, were used for cross-validation.

Table 1. Potential means of contact tracing for passengers from the Diamond Princess cruise ship during the 1-day Taiwan tour.

Digital records	Investigation	Difficulties
Travelling guide provided by the agency	To trace the travelling routes	• Cannot trace those travelling with taxies or independent of the agency • Cannot identify the exact visiting time
GPS in buses	To trace locations where buses drove through	• Cannot trace those travelling with taxies or independent of the agency
Transaction of credit cards	To trace the travelling routes by shopping records	• Not all passengers used credit cards
CCTV[a] and recognition system of vehicle license plates	To trace the travelling routes	• Large number of passengers • Depends on CCTV location • Time consumption
Electronic toll collection system on the nationwide freeways	To trace locations where buses drove through	• Available for limited routes
Mobile phone positioning system	To trace individual-based travelling route	• Only applicable to those using roaming service for mobile phone

[a]CCTV: closed-circuit television.

Routes of Passengers from the Diamond Princess Cruise Ship and Contact Tracing

Figure 1 shows the travelers' routes while passengers got off the cruise. Based on the estimated routes, the 39 marked locations were identified on the map list as shown in Figure 2. Most of the warning locations are famous sight-seeing visiting areas in northern Taiwan.

Figure 1. Passengers' routes of the 1-day tour on January 31, 2020, in Taiwan.

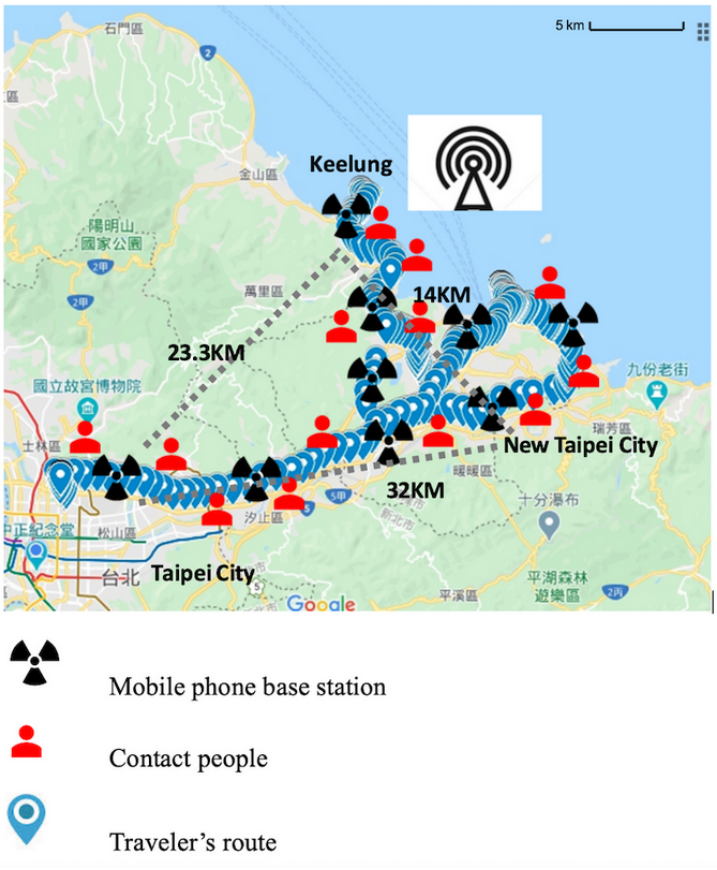
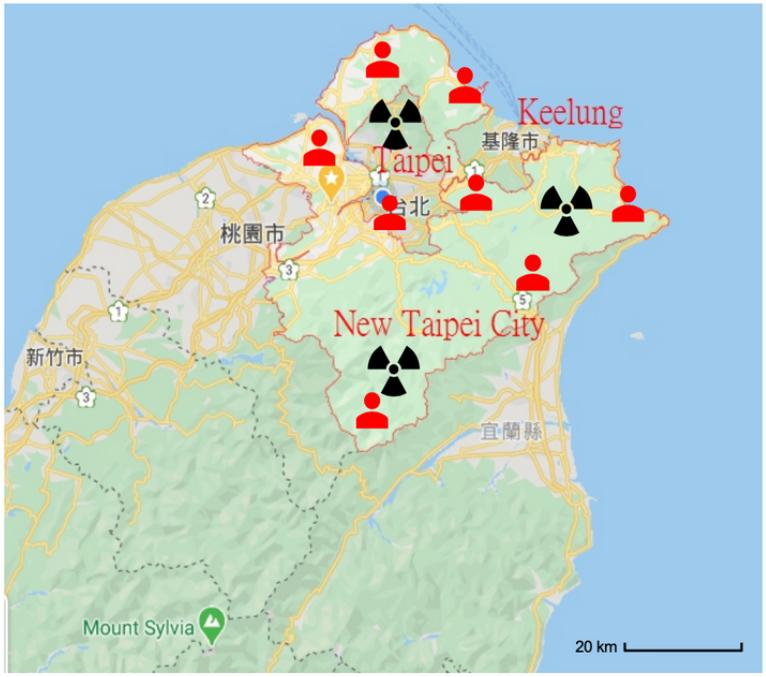


Figure 2. The tagged locations based on routes of passengers in three cities (Keelung, New Taipei City, and Taipei) for broadcasting the warning message. Tagged locations: (1) Keelung: Keelung Waimu Mountain, Keelung Zhongzheng Park, Keelung Cultural Center, Xiandong, Keelung, Zhengbin Fishing Port, Keelung Peace Island, Keelung Harbor, Miaokou Night Market, White Rice Fortress, Keelung, Diaohe St., Zhongzheng Dist., Keelung City. (2) New Taipei City: Shifen Old Street, Jiufen, Shifen Waterfall, New Taipei City, Yehliu Geopark, NanYa Wonderful Rock, Ruifang District, New Taipei City, Turtle Roar, Wanli District, Bisha Fishing Harbor. (3) Taipei City: Nanmen Market, National Revolutionary Martyrs' Shrine, Jhungshan Auditorium, National Chiang Kai-shek Memorial Hall, Taipei, Confucius Temple, Liberty Square Taipei, Ximending, Section 4, Zhongxiao East Road, Taipei City, Dalongdong Bao'an Temple, National Palace Museum, Dihua Street, Zhuzihu, Yangmingshan, The Grand Hotel Taipei, Xichang Street: Herb Lane, Lungshan Temple, Taipei Main Station, Pacific Sogo(Fuxing), Taipei.



Based on the mobile position information, we identified 627,386 corresponding possible contact-persons. The symptom monitoring and self-quarantine message was sent though SMS after identifying the contact-person on February 7, 2020. The alert message (Figure 3A) was as follows: "Due to the COVID-19 epidemic, anyone who had been to the following locations from 6 a.m. to 5:30 p.m. on Jan. 31, please conduct symptom monitoring and self-quarantine and isolation until Feb. 14. If you need any assistance please call 1922 hotline."

The specification of the locations can be revealed as shown in Figure 3B [6] via SMS. As of the end of March 2020, it has been visited 29,317,172 times.

Figure 3. A demonstration of the (A) alert message and (B) contact locations integrated with Google Maps.

RT-PCR Test for People with Suspected Symptoms and Signs Notified Through The Alert Message

As of February 14, the CECC was notified of 20 symptomatic contacts who followed the guidance on the alert message and thus sought medical care. All received COVID-19 testing and all tested negative.

COVID-19 Surveillance for the Contact Population Through National Health Insurance Claims Data

As of February 29, a total of 121 hospitalizations due to pneumonia were identified among the contact population. A total of 24 contact-patients had been reported as suspected COVID-19 cases and all tested negative. Among the 41 contact-patients who remained hospitalized, 23 received testing for COVID-19 and all showed negative results.

Disease Surveillance of Respiratory Syndrome and Pneumonia for the Contact Population Through Big Health Insurance Claimed Data

During the surveillance period between January 31 and March 10, 2020, the age-standardized rate of respiratory syndrome (16.87 per 1000) in the contact population was lower than that in the general population (19.23 per 1000; Table 2). We noted an age-standardized relative ratio of 0.929 (95% CI 0.923-0.935) for the contact population diagnosed as having respiratory syndrome, relative to the general population. Similarly, there was a lower risk for pneumonia among the contact population compared with the general population (age-standardized relative ratio 0.915, 95% CI 0.869-0.963). This suggests that smart contract tracing with mobile position data followed by self-quarantine and isolation may be a useful means of preventing the spread of COVID-19.

Table 2. Respiratory syndrome/pneumonia cases, rates, and relative ratios in the exposed group and unexposed groups.

Diseases	Exposed group (n=627,386)	Unexposed group (n=23,877,447)
Respiratory syndrome		
Cases, n (%)	105,837 (16.87)	4,592,694 (19.23)
Expected cases, n (%)	113,920 (18.16)	N/A[a]
Crude relative ratio (95% CI)	0.877 (0.872-0.882)	1.00
Age-standardized relative ratio (95% CI)	0.929 (0.923-0.935)	1.00
Pneumonia		
Cases, n (%)	1,479 (0.236)	91,066 (0.381)
Expected cases, n (%)	1,616 (0.26)	N/A
Crude relative ratio (95% CI)	0.618 (0.587-0.651)	1.00
Age-standardized relative ratio (95% CI)	0.915 (0.869-0.963)	1.00

[a]N/A: not applicable.

Discussion

Although the public health interventions aimed at reducing the population contact rates have demonstrated their efficacy in containing the pandemic [7,8], their implementation has a great impact for subjects, the community, and the public health system in a contemporary democratic society [9]. The manpower and workload for quarantine of infectious disease are usually highly demanding. When facing the global crisis of an emerging infectious disease such as COVID-19, rapid response and immediate interventions for preventing the outbreak are essential. Smart contact tracing with big sensor data on mobile position data and its connected mobile phone can provide information in a timely manner and help crisis management under such a situation. In this study, we demonstrated how smart contact tracing can be applied to the contact history between Taiwanese people and the possible infected passengers who disembarked from the Diamond Princess cruise ship just before the outbreak of COVID-19 on the cruise ship.

In order to rapidly trace potential contacts, numerous locations where the cruise ship passengers may have visited were first identified by using passive mobile positioning data. These data handovers in network cells of mobile service providers stored the location of call activities. These data have good potential for not only monitoring the mobility of the tourist group but also identifying people in contact with the tour group. This mobile geopositioning method had also been used in a mobile health study to measure human mobility, disease connectivity, and health risk in travelers [10].

Based on our mobility and geography of mobile position analysis, a total of 627,386 citizens were possibly exposed to passengers on the Diamond Princess cruise. These persons were sent syndrome monitoring and self-quarantine information via SMS messaging to their phones for mitigating the possible community spread. Although over 190 contact persons per traveler might not be realistic, increasing the targeted contact population with no harm was acceptable as a step against COVID-19 spread in this emergency situation. Moreover, it should be noted that providing accurate, timely information for decision making is crucial during the crisis. It had been a few days since the exposure of infected hosts had occurred. Contact tracing and management using information technology had to be quickly implemented without delay. This is one of the advantages of using big data technology for analysis. In terms of fighting against COVID-19, similar big data technologies have been applied in the spatial tracking of patients for tracking virus transmission and potential spatiotemporal exposure, to support the epidemiological investigation with rapid analysis [11].

On evaluating the impact of self-quarantine at home policy with an alert message, no confirmed COVID-19 cases in this contact population were ascertained. In addition, we used the National Health Insurance claimed database to facilitate another big data analysis for the surveillance of severe respiratory symptoms. The lower risk of mild or severe respiratory symptoms was noted in the exposed group (contact population) compared with the unexposed group. In addition to the prevention of the spread of SARS-CoV-2, this may be attributed to the enhanced awareness of a patient's own health status and the enhanced personal self-contained lifestyle affected by the alert message.

From the perspective of big data technology, this study identified the mobile geopositioning method through big data technology as an effective method for achieving geographic route acquisition and mapping mobile positioning from mobility of the population nearby, to conduct contact tracing. The analysis platform was quickly constructed through an innovative technology system to support timely epidemic analysis.

More importantly, smart contact tracing with big sensor data analysis applied to contact investigation of those who may have contracted COVID-19 is also cost-effective because the costs and manpower would be substantially reduced compared to use of the conventional epidemiological contact investigation method.

There is a major limitation to our methods. The potential contact-persons identified in this study tended to include more working populations and students because those who have no mobile phone are more likely to be very young or elderly people who are hard to be traced by this smart contact tracing

technology. This weakness may be tackled by providing an active surveillance system for them to contact local health authority.

In conclusion, this study demonstrated the successful prevention of community spread of COVID-19 in the crisis of contact from potential infected travelers of the Diamond Princess cruise ship by using big data analytic. This is an example of how big data technology can be applied in contract tracing and quarantine to support the epidemiological surveillance of new virus infection.

Acknowledgments

This work was funded by the “Innovation and Policy Center for Population Health and Sustainable Environment (Population Health Research Center, PHRC), College of Public Health, National Taiwan University” from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

Authors' Contributions

C-MC and C-CC contributed to the study concept and design. H-WJ, S-CC, P-CL, C-FL, Y-TY, and M-YC contributed to the acquisition of data. H-HJ and C-YH contributed to the statistical analysis. L-SC and H-HC contributed to the interpretation of results. C-MC drafted the manuscript. All authors gave final approval for the manuscript.

Conflicts of Interest

None declared.

References

1. World Health Organization. 2020. Coronavirus disease (COVID-2019) situation reports URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/ [accessed 2020-05-04]
2. Cheng H, Li S, Yang C. Initial rapid and proactive response for the COVID-19 outbreak - Taiwan's experience. J Formos Med Assoc 2020 Apr;119(4):771-773 [FREE Full text] [doi: 10.1016/j.jfma.2020.03.007] [Medline: 32222336]
3. Park S, Choi GJ, Ko H. Information Technology-Based Tracing Strategy in Response to COVID-19 in South Korea-Privacy Controversies. JAMA 2020 Apr 23. [doi: 10.1001/jama.2020.6602] [Medline: 32324202]
4. Wang CJ, Ng CY, Brook RH. Response to COVID-19 in Taiwan: Big Data Analytics, New Technology, and Proactive Testing. JAMA 2020 Mar 03;323(14):1341. [doi: 10.1001/jama.2020.3151] [Medline: 32125371]
5. Breslow N, Day N. Statistical methods in cancer research. Volume II--The design and analysis of cohort studies. IARC Sci Publ 1987(82):1-406. [Medline: 3329634]
6. Google Maps. URL: http://bit.ly/2SpSxeT [accessed 2020-05-04]
7. Ferguson NM, Cummings DA, Cauchemez S, Fraser C, Riley S, Meeyai A, et al. Strategies for containing an emerging influenza pandemic in Southeast Asia. Nature 2005 Sep 08;437(7056):209-214. [doi: 10.1038/nature04017] [Medline: 16079797]
8. Longini IM, Nizam A, Xu S, Ungchusak K, Hanshaoworakul W, Cummings DAT, et al. Containing pandemic influenza at the source. Science 2005 Aug 12;309(5737):1083-1087 [FREE Full text] [doi: 10.1126/science.1115717] [Medline: 16079251]
9. Fisher D, Wilder-Smith A. The global community needs to swiftly ramp up the response to contain COVID-19. The Lancet 2020 Apr;395(10230):1109-1110. [doi: 10.1016/s0140-6736(20)30679-6]
10. Lai S, Farnham A, Ruktanonchai N, Tatem A. Measuring mobility, disease connectivity and individual risk: a review of using mobile phone data and mHealth for travel medicine. J Travel Med 2019 May 10;26(3):a [FREE Full text] [doi: 10.1093/jtm/taz019] [Medline: 30869148]
11. Zhou C, Su F, Pei T, Zhang A, Du Y, Luo B, et al. COVID-19: Challenges to GIS with Big Data. Geography and Sustainability 2020 Mar. [doi: 10.1016/j.geosus.2020.03.005]

Abbreviations

CCTV: closed-circuit television
CECC: Central Epidemic Command Center
COVID-19: coronavirus disease
ICD-10: The International Statistical Classification of Diseases and Related Health Problems, 10th Revision
SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2
WHO: World Health Organization

Edited by G Eysenbach; submitted 22.04.20; peer-reviewed by H Sjursen, A Benis; accepted 30.04.20; published 05.05.20.

Please cite as:
Chen CM, Jyan HW, Chien SC, Jen HH, Hsu CY, Lee PC, Lee CF, Yang YT, Chen MY, Chen LS, Chen HH, Chan CC
Containing COVID-19 Among 627,386 Persons in Contact With the Diamond Princess Cruise Ship Passengers Who Disembarked in Taiwan: Big Data Analytics
J Med Internet Res 2020;22(5):e19540
URL: http://www.jmir.org/2020/5/e19540/
doi:10.2196/19540
PMID:32353827

Original Paper

Creating COVID-19 Stigma by Referencing the Novel Coronavirus as the "Chinese virus" on Twitter: Quantitative Analysis of Social Media Data

Henna Budhwani[1], MPH, PhD; Ruoyan Sun[1], MHS, PhD

Department of Health Care Organization and Policy, School of Public Health, University of Alabama at Birmingham, Birmingham, AL, United States

Corresponding Author:
Henna Budhwani, MPH, PhD
Department of Health Care Organization and Policy
School of Public Health
University of Alabama at Birmingham
1720 University Blvd
RPHB #330C
Birmingham, AL, 35294
United States
Phone: 1 2059757613
Email: budhwani@uab.edu

Abstract

Background: Stigma is the deleterious, structural force that devalues members of groups that hold undesirable characteristics. Since stigma is created and reinforced by society—through in-person and online social interactions—referencing the novel coronavirus as the "Chinese virus" or "China virus" has the potential to create and perpetuate stigma.

Objective: The aim of this study was to assess if there was an increase in the prevalence and frequency of the phrases "Chinese virus" and "China virus" on Twitter after the March 16, 2020, US presidential reference of this term.

Methods: Using the Sysomos software (Sysomos, Inc), we extracted tweets from the United States using a list of keywords that were derivatives of "Chinese virus." We compared tweets at the national and state levels posted between March 9 and March 15 (preperiod) with those posted between March 19 and March 25 (postperiod). We used Stata 16 (StataCorp) for quantitative analysis, and Python (Python Software Foundation) to plot a state-level heat map.

Results: A total of 16,535 "Chinese virus" or "China virus" tweets were identified in the preperiod, and 177,327 tweets were identified in the postperiod, illustrating a nearly ten-fold increase at the national level. All 50 states witnessed an increase in the number of tweets exclusively mentioning "Chinese virus" or "China virus" instead of coronavirus disease (COVID-19) or coronavirus. On average, 0.38 tweets referencing "Chinese virus" or "China virus" were posted per 10,000 people at the state level in the preperiod, and 4.08 of these stigmatizing tweets were posted in the postperiod, also indicating a ten-fold increase. The 5 states with the highest number of postperiod "Chinese virus" tweets were Pennsylvania (n=5249), New York (n=11,754), Florida (n=13,070), Texas (n=14,861), and California (n=19,442). Adjusting for population size, the 5 states with the highest prevalence of postperiod "Chinese virus" tweets were Arizona (5.85), New York (6.04), Florida (6.09), Nevada (7.72), and Wyoming (8.76). The 5 states with the largest increase in pre- to postperiod "Chinese virus" tweets were Kansas (n=697/58, 1202%), South Dakota (n=185/15, 1233%), Mississippi (n=749/54, 1387%), New Hampshire (n=582/41, 1420%), and Idaho (n=670/46, 1457%).

Conclusions: The rise in tweets referencing "Chinese virus" or "China virus," along with the content of these tweets, indicate that knowledge translation may be occurring online and COVID-19 stigma is likely being perpetuated on Twitter.

(J Med Internet Res 2020;22(5):e19301) doi:10.2196/19301

KEYWORDS

COVID-19; coronavirus; Twitter; stigma; social media; public health

Introduction

Stigma is the deleterious, structural force that devalues those who hold undesirable characteristics [1]. Stigma is a social process that occurs between groups; this process can occur in-person and online [2-6]. Regardless of setting, research has consistently found that stigma is associated with negative health outcomes [2,4,6-9]. For example, HIV-related stigma has pushed the HIV-epidemic underground, fueling ongoing transmission [10], and other disease-related stigmas are associated with negative health outcomes ranging from missed clinical visits to suicidal ideation [1,6,9]. There is evidence to show that stigma can become internalized, and internalized stigma can lead to distrust of health professionals, skepticism of public health systems, and an unwillingness to disclose behaviors related to transmission [2,8,9]. Because the coronavirus disease (COVID-19) is infectious, contact tracing is critically important to assessing community spread; thus, it is imperative that individuals trust their public health and health care systems so that they are willing to accept testing and, if diagnosed with COVD-19, report their whereabouts and activities. Therefore, creating and perpetuating stigma related to COVID-19 could be detrimental to public health efforts that require potentially stigmatized individuals to engage with their health systems.

On March 16, 2020, the president of the United States referred to the novel coronavirus as the "Chinese virus" on Twitter. He tweeted "The United States will be powerfully supporting those industries... that are particularly affected by the Chinese Virus..." After this presidential reference, a dialogue emerged examining if the phrase "Chinese virus" was xenophobic and stigmatizing, considering the availability of alternative scientific names such as coronavirus or COVID-19. Since stigma is created and perpetuated by society through social interaction and public commentary (eg, use of the term "Chinese virus" instead of scientific terms on Twitter), and stigma is reinforced by those in power (eg, use of the term "Chinese virus" by the US president), we hypothesized that there would be an increase in the frequency of the phrases "Chinese virus" and "China virus" on Twitter, comparing the prevalence of these phrases before and after the presidential reference.

Methods

Twitter

Twitter is an online social media platform where users send and receive short posts (maximum 280 characters) called tweets. Twitter currently has 152 million daily users, who produce about 500 million daily tweets [11].

Data, Tweets

We downloaded tweets from all 50 US states, using the Sysomos software (Sysomos, Inc). We extracted tweets that mentioned "Chinese virus" or "China virus" but did not contain "COVID-19" or "coronavirus." The list of keywords referencing the "Chinese virus" are "Chinesevirus," "Chinese virus," "Chinavirus," "China virus," "#ChineseVirus19," "#Chinesevirus," "#ChineseVirusCorona," and "#Chinavirus." We excluded tweets containing the keywords "coronavirus," "corona virus," "COVID-19," "COVID19," "#COVID2019," and "#corona." By excluding tweets that contained both "Chinese virus" and "coronavirus," we collated a sample of tweets that represented the intent of using "Chinese virus" in place of a scientific alternative, likely indicating deliberate stigmatization. We imputed the location of tweets based on Twitter users' self-reported state of residence. Tweets posted between March 9 and March 15, 2020 (preperiod), were compared with tweets posted between March 19 and March 25, 2020 (postperiod). Original tweets and quote tweets (adding comments to an existing tweet) were included but not retweets (reposting of an existing tweet). Our final sample (N=193,862) contained all tweets posted in the pre- and postperiods by US-based Twitter users that exclusively mentioned a derivative of "Chinese virus." Data extraction was conducted on April 10, 2020. Ethical approval was provided by the University of Alabama at Birmingham Institutional Review Board (IRB-#300005071).

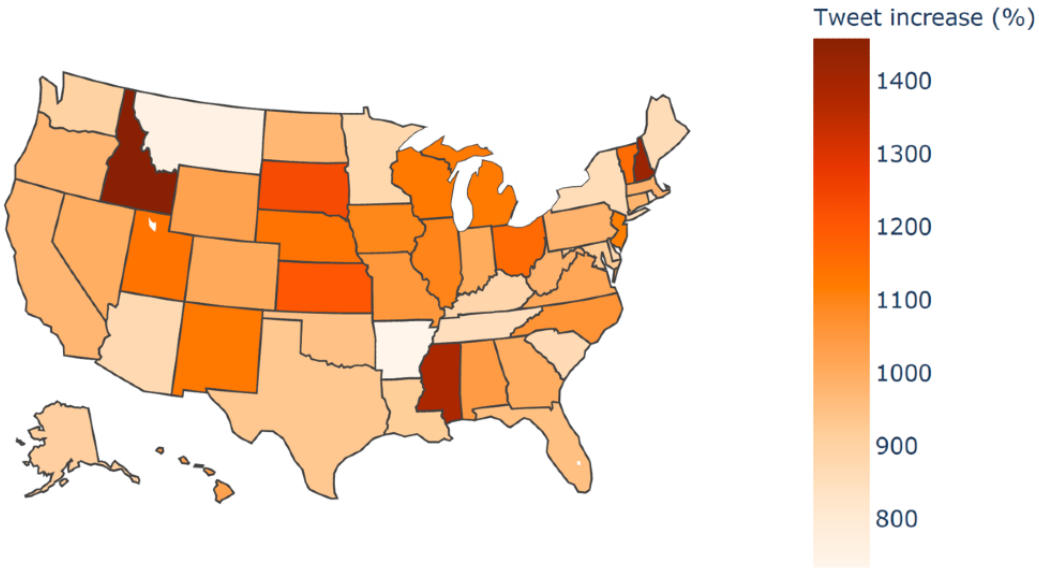
Analysis

We used Stata 16 (StataCorp) to analyze our Twitter data and Python software (Python Software Foundation) to plot our state-level gradient heat map.

Results

A total of 16,535 "Chinese virus" or "China virus" tweets were identified in the preperiod, and 177,327 tweets were identified in the postperiod, illustrating a 972.43% (n=160,792/16,535) increase. Comparatively, the number of tweets referencing COVID-19 in the preperiod and postperiod remained steady, at about 4.9 million tweets per period. A total of 13,569 (82.06%) of the preperiod and 145,521 (82.06%) of the postperiod tweets were associated with a Twitter user's self-reported US state. Figure 1 is a heat map illustrating the state-by-state increases of tweets referencing "Chinese virus" or "China virus." The darker the shade, the greater the increase. All 50 US states witnessed an increase in the number of tweets exclusively mentioning "Chinese virus" or "China virus" rather than COVID-19 or coronavirus. The 5 US states with the highest number of postperiod "Chinese virus" tweets were Pennsylvania, New York, Florida, Texas, and California. The 5 US states with the largest increase in pre- to postperiod "Chinese virus" tweets were Kansas, South Dakota, Mississippi, New Hampshire, and Idaho.

Figure 1. Heat map of increases in tweets referencing “Chinese virus” or “China virus” across the United States.



In Table 1, we present US state-level results of tweets referencing “Chinese virus” or “China virus.” On average, at the state level, 271 such tweets were found in the preperiod and 2910 in the postperiod, indicating a ten-fold increase, similar to what we found at the national level. We also calculated the percentage increase and the prevalence increase. The percentage increase measures the percentage of all COVID-19 related tweets that mentioned “China virus” or “Chinese virus” exclusively. To account for variations in population size, prevalence of “Chinese virus” tweets per 10,000 people for each US state was calculated using the following formula: . State population sizes were taken from the 2019 US Census Bureau estimates [12]. On average, the state-level percentage increase was 997%, with a minimum of 661% and a maximum of 1447%. Similarly, the prevalence increase mean was 1015%, with a minimum of 734% and a maximum of 1456%. Large variations were found across US states, with the lowest postperiod prevalence of “Chinese virus” or “China virus” in South Dakota and the highest in Wyoming. The 5 US states with the highest prevalence of “Chinese virus” or “China virus” postperiod tweets were Arizona, New York, Florida, Nevada, and Wyoming.

Table 1. Tweets referencing the novel coronavirus as "Chinese virus" or "China virus" by state.

States	Preperiod				Postperiod				Change from pre- to postperiod	
	COVID-19 tweets, n	"Chinese virus" tweets, n	Percentage of tweets[a], (%)	Prevalence of tweets[b]	COVID-19 tweets, n	"Chinese virus" tweets, n	Percentage of tweets[a], (%)	Prevalence of tweets[b]	Percentage increase[c] (%)	Prevalence increase[d] (%)
AL	40,588	153	0.38	0.31	39,434	1749	4.44	3.57	1077	1043
AK	9251	40	0.43	0.55	9597	404	4.21	5.52	874	910
AZ	83,019	438	0.53	0.60	89,127	4256	4.78	5.85	805	872
AR	21,810	109	0.50	0.36	22,741	910	4.00	3.02	701	735
CA	696,645	1806	0.26	0.46	685,596	19,442	2.84	4.92	994	977
CO	84,092	291	0.35	0.51	85,014	3218	3.79	5.59	994	1006
CT	40,304	116	0.29	0.33	40,531	1253	3.09	3.51	974	980
DE	9789	31	0.32	0.32	10,095	304	3.01	3.12	851	881
FL	270,723	1243	0.46	0.58	294,652	13,070	4.44	6.09	866	951
GA	135,543	382	0.28	0.36	136,875	4192	3.06	3.95	987	997
HI	15,261	53	0.35	0.37	18,237	597	3.27	4.22	843	1026
ID	13,810	46	0.33	0.26	14,683	716	4.88	4.01	1364	1457
IL	176,425	410	0.23	0.32	169,849	4918	2.90	3.88	1146	1100
IN	58,767	192	0.33	0.29	57,218	2118	3.70	3.15	1033	1003
IA	27,552	71	0.26	0.23	27,917	847	3.03	2.68	1077	1093
KS	24,678	58	0.24	0.20	24,694	755	0.31	2.59	1201	1202
KY	45,648	179	0.39	0.40	45,841	1765	3.85	3.95	882	886
LA	51,734	151	0.29	0.32	48,623	1535	3.16	3.30	982	917
ME	16,948	54	0.32	0.40	17,762	520	2.93	3.87	819	863
MD	75,527	189	0.25	0.31	76,274	1932	2.53	3.20	912	922
MA	138,665	295	0.21	0.43	137,279	3201	2.33	4.64	996	985
MI	108,514	297	0.27	0.30	103,934	3623	3.49	3.63	1174	1120
MN	63,304	192	0.30	0.34	65,570	1882	2.87	3.34	846	880
MS	19,530	54	0.28	0.18	18,771	803	4.28	2.70	1447	1387
MO	68,869	201	0.29	0.33	71,951	2317	3.22	3.78	1003	1053
MT	9365	61	0.65	0.57	10,503	521	4.96	4.87	662	754
NE	19,791	54	0.27	0.28	18,840	670	3.56	3.46	1203	1141
NV	52,996	217	0.41	0.70	53,730	2377	4.42	7.72	980	995
NH	14,260	41	0.29	0.30	15,096	623	4.13	4.58	1335	1420
NJ	96,806	315	0.33	0.35	100,334	3823	3.81	4.30	1071	1114
NM	18,966	51	0.27	0.24	20,220	627	3.10	2.99	1053	1129
NY	487,901	1225	0.25	0.63	484,515	11,754	2.43	6.04	866	860
NC	110,832	327	0.30	0.31	115,394	3795	3.29	3.62	1015	1061
ND	5649	18	0.32	0.24	6148	193	3.14	2.53	885	972
OH	145,371	366	0.25	0.31	127,421	4613	3.62	3.95	1338	1160
OK	33,480	137	0.41	0.35	33,857	1436	4.24	3.63	937	948
OR	64,817	185	0.29	0.44	65,972	1985	3.01	4.71	954	973
PA	159,712	485	0.30	0.38	161,156	5249	3.26	4.10	973	982
RI	14,234	43	0.30	0.41	14,219	385	2.71	3.63	796	795

States	Preperiod				Postperiod				Change from pre- to postperiod	
	COVID-19 tweets, n	"Chinese virus" tweets, n	Percentage of tweets[a], (%)	Prevalence of tweets[b]	COVID-19 tweets, n	"Chinese virus" tweets, n	Percentage of tweets[a], (%)	Prevalence of tweets[b]	Percentage increase[c] (%)	Prevalence increase[d] (%)
SC	43,104	222	0.52	0.43	46,251	2145	4.64	4.17	800	866
SD	6252	15	0.24	0.17	6573	200	3.04	2.26	1168	1233
TN	82,478	361	0.44	0.53	82,050	3431	4.18	5.02	855	850
TX	378,047	1442	0.38	0.50	369,006	14,861	4.03	5.13	956	931
UT	30,422	81	0.27	0.25	28,464	1004	3.53	3.13	1225	1140
VT	8625	18	0.21	0.29	9527	226	2.37	3.62	1037	1156
VA	97,602	301	0.31	0.35	104,176	3351	3.22	3.93	943	1013
WA	123,025	331	0.27	0.43	116,656	3316	2.84	4.35	957	902
WV	15,523	47	0.30	0.26	15,698	509	3.24	2.84	971	983
WI	51,670	130	0.25	0.22	52315	1593	3.05	2.74	1110	1125
WY	6185	45	0.73	0.78	6875	507	7.37	8.76	914	1027
Mean	87,482	271	0.33	0.38	87,545	2910	3.57	4.08	997	1015

[a]Percentage of all COVID-19 related tweets that mentioned "Chinese virus" or "China virus" exclusively.

[b]Prevalence of "Chinese virus" tweets per 10,000 people was calculated using the following formula:  .

[c]Percentage of increase was calculated as:  .

[d]Prevalence increase was calculated as:  .

Discussion

Principal Result

We found notable increases in the use of the terms "Chinese virus" and "China virus" on Twitter at both the national and state levels by comparing these tweets (percentage and prevalence) both before and after the March 16, 2020, presidential reference. The following are examples of "Chinese virus" or "China virus" tweets:

- Not parroting MSM's [main stream media's] narrative. It's the #WuFlu #ChineseCoronaVirus #ChinaVirus"
- "#ChinaVirus #ChinaLiesPeopleDie"

Limitations

The pandemic is currently underway, so Twitter data—both in quantity (quantitative) and content (qualitative)—are rapidly shifting. We were unable to screen for automatically generated tweets (bots) within this short report [13,14]. Geographic locations associated with Twitter accounts were self-reported; thus, it is possible that some Twitter users may have moved without updating their state location or may have reported a false state location.

Comparison With Prior Work

There is a growing body of academic literature that leverages Twitter data to assess trends in population health and public sentiment [15-17]. Chew and Eysenbach [18] conducted a seminal examination of knowledge translation using Twitter data during the H1N1 outbreak; they found the proportion of tweets using "H1N1" increased over time compared to the relative use of "swine flu," suggesting that the media's choice in terminology (shifting from using the term "swine flu" to "H1N1") influenced public uptake. In addition, it is relevant that a recent publication by Logie and Turan [19] presented a narrative on how stigma can hurt the COVID-19 public health response. This short report was developed considering the findings from prior studies.

Future Research

Future research could evaluate and show that stigma mechanisms work online, validate if Twitter and social media data can be informative to epidemic surveillance and health communication, examine the extent that Twitter and social media data is reliable in informing public health efforts and social science research, and explore how Twitter users view COVID-19 and the COVID-19 public health response (eg, testing, linkage to care).

Additionally, although there is a growing body of research using tweets to examine aspects of the novel coronavirus [20-22], to our knowledge, no studies have included a comprehensive set of search terms, which may include phrases such as "ncov," "covid," "sars-cov," and "rona," in defining their samples. If data extraction is not comprehensive, we run the risk of missing emerging sentiments and terminology, such as referencing the novel coronavirus as the "China virus" or "Chinese virus," and sociobehavioral outcomes related to these trends.

Conclusions

The rise in tweets citing "Chinese virus" or "China virus" instead of COVID-19 or the novel coronavirus after the presidential reference on Twitter, along with the content of these

tweets, indicate that knowledge translation may be occurring online and COVID-19 stigma is likely being perpetuated on Twitter. Generally speaking, perpetuating COVID-19-related stigma by using the phrase "Chinese virus" could harm public health efforts related to addressing the pandemic, specifically inciting fear and increasing distrust of public health systems by Chinese and Asian Americans. If these stigmatizing terms persist as malicious synonyms for the novel coronavirus, reparative efforts may be required to restore trust by marginalized communities.

Acknowledgments

Research reported in this publication was supported by the University of Alabama at Birmingham School of Public Health Back of the Envelope (for RS) and the National Institute of Mental Health of the National Institutes of Health under Award Number 1K01MH116737 (for HB). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Authors' Contributions

HB conceptualized this study, and RS conducted the data collection and analysis. Both authors contributed to manuscript development and writing.

Conflicts of Interest

None declared.

References

1. Goffman E. Stigma: Notes on the Management of Spoiled Identity. Upper Saddle River, NJ: Prentice Hall; 1963.
2. Budhwani H, De P. Perceived stigma in health care settings and the physical and mental health of people of color in the United States. Health Equity 2019;3(1):73-80 [FREE Full text] [doi: 10.1089/heq.2018.0079] [Medline: 30915422]
3. Ho CL, Pan W, Taylor LD. Stigma of HIV testing on online HIV forums: self-stigma and the unspoken. J Psychosoc Nurs Ment Health Serv 2017 Dec 01;55(12):34-43. [doi: 10.3928/02793695-20170905-01] [Medline: 28892555]
4. Karamouzian M, Knight R, Davis WM, Gilbert M, Shoveller J. Stigma associated with sexually transmissible infection testing in an online testing environment: examining the perspectives of youth in Vancouver, Canada. Sex Health 2018;15(1):46. [doi: 10.1071/sh17089]
5. Milin R, Kutcher S, Lewis SP, Walker S, Wei Y, Ferrill N, et al. Impact of a mental health curriculum on knowledge and stigma among high school students: a randomized controlled trial. J Am Acad Child Adolesc Psychiatry 2016 May;55(5):383-391.e1. [doi: 10.1016/j.jaac.2016.02.018] [Medline: 27126852]
6. Pachankis JE, Hatzenbuehler ML, Wang K, Burton CL, Crawford FW, Phelan JC, et al. The burden of stigma on health and well-being: a taxonomy of concealment, course, disruptiveness, aesthetics, origin, and peril across 93 stigmas. Pers Soc Psychol Bull 2018 Apr;44(4):451-474 [FREE Full text] [doi: 10.1177/0146167217741313] [Medline: 29290150]
7. Budenz A, Klassen A, Purtle J, Yom Tov E, Yudell M, Massey P. Mental illness and bipolar disorder on Twitter: implications for stigma and social support. J Ment Health 2020 Apr;29(2):191-199. [doi: 10.1080/09638237.2019.1677878] [Medline: 31694433]
8. Budhwani H, Hearld KR, Milner AN, Charow R, McGlaughlin EM, Rodriguez-Lauzurique M, et al. Transgender women's experiences with stigma, trauma, and attempted suicide in the Dominican Republic. Suicide Life Threat Behav 2018 Dec;48(6):788-796. [doi: 10.1111/sltb.12400] [Medline: 28950402]
9. Turan B, Budhwani H, Fazeli PL, Browning WR, Raper JL, Mugavero MJ, et al. How does stigma affect people living with HIV? The mediating roles of internalized and anticipated HIV stigma in the effects of perceived community stigma on health and psychosocial outcomes. AIDS Behav 2017 Jan;21(1):283-291 [FREE Full text] [doi: 10.1007/s10461-016-1451-5] [Medline: 27272742]
10. Pachankis JE, Hatzenbuehler ML, Hickson F, Weatherburn P, Berg RC, Marcus U, et al. Hidden from health. AIDS 2015;29(10):1239-1246. [doi: 10.1097/qad.0000000000000724]
11. Dai H, Deem MJ, Hao J. Geographic variations in electronic cigarette advertisements on Twitter in the United States. Int J Public Health 2017 May;62(4):479-487. [doi: 10.1007/s00038-016-0906-9] [Medline: 27742923]
12. United States Census Bureau. 2019 Dec. Table 1. Annual Estimates of the Resident Population for the United States, Regions, States, and Puerto Rico: April 1, 2010 to July 1, 2019 (NST-EST2019-01) URL: https://www.census.gov/newsroom/press-kits/2019/national-state-estimates.html [accessed 2020-04-22]
13. Allem J, Ferrara E. The importance of debiasing social media data to better understand e-cigarette-related attitudes and behaviors. J Med Internet Res 2016 Aug 09;18(8):e219. [doi: 10.2196/jmir.6185] [Medline: 27507563]
14. Allem J, Ferrara E, Uppu SP, Cruz TB, Unger JB. E-cigarette surveillance with social media data: social bots, emerging topics, and trends. JMIR Public Health Surveill 2017 Dec 20;3(4):e98. [doi: 10.2196/publichealth.8641] [Medline: 29263018]

15. Alessa A, Faezipour M. Flu outbreak prediction using Twitter posts classification and linear regression with historical Centers for Disease Control and Prevention reports: prediction framework study. JMIR Public Health Surveill 2019 Jun 25;5(2):e12383. [doi: 10.2196/12383] [Medline: 31237567]
16. Grajales FJ, Sheps S, Ho K, Novak-Lauscher H, Eysenbach G. Social media: a review and tutorial of applications in medicine and health care. J Med Internet Res 2014 Feb 11;16(2):e13. [doi: 10.2196/jmir.2912] [Medline: 24518354]
17. Ji X, Chun SA, Wei Z, Geller J. Twitter sentiment classification for measuring public health concerns. Soc Netw Anal Min 2015;5(1):13 [FREE Full text] [doi: 10.1007/s13278-015-0253-5] [Medline: 32226558]
18. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118. [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
19. Logie CH, Turan JM. How do we balance tensions between COVID-19 public health responses and stigma mitigation? learning from HIV research. AIDS Behav 2020 Apr 07:e. [doi: 10.1007/s10461-020-02856-8] [Medline: 32266502]
20. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of Tweeters during the COVID-19 pandemic: infoveillance study. J Med Internet Res 2020 Apr 21;22(4):e19016. [doi: 10.2196/19016] [Medline: 32287039]
21. Kouzy R, Abi Jaoude J, Kraitem A, El Alam MB, Karam B, Adib E, et al. Coronavirus goes viral: quantifying the COVID-19 misinformation epidemic on Twitter. Cureus 2020 Mar 13;12(3):e7255 [FREE Full text] [doi: 10.7759/cureus.7255] [Medline: 32292669]
22. Rosenberg H, Syed S, Rezaie S. The Twitter pandemic: the critical role of Twitter in the dissemination of medical information and misinformation during the COVID-19 pandemic. CJEM 2020 Apr 06:1-4 [FREE Full text] [doi: 10.1017/cem.2020.361] [Medline: 32248871]

Abbreviations

COVID-19: coronavirus disease

Edited by G Eysenbach; submitted 12.04.20; peer-reviewed by E Da Silva, JP Allem; comments to author 21.04.20; revised version received 23.04.20; accepted 26.04.20; published 06.05.20.

Please cite as:
Budhwani H, Sun R
Creating COVID-19 Stigma by Referencing the Novel Coronavirus as the "Chinese virus" on Twitter: Quantitative Analysis of Social Media Data
J Med Internet Res 2020;22(5):e19301
URL: http://www.jmir.org/2020/5/e19301/
doi:10.2196/19301
PMID:32343669

Original Paper

Impact of Online Information on Self-Isolation Intention During the COVID-19 Pandemic: Cross-Sectional Study

Ali Farooq[1], MCS, MBA, MSc, DSc; Samuli Laato[1], MSc; A K M Najmul Islam[1], PhD

Department of Future Technologies, University of Turku, Turku, Finland

Corresponding Author:
Ali Farooq, MCS, MBA, MSc, DSc
Department of Future Technologies
University of Turku
Vesilinnantie 3
Turku, 20500
Finland
Phone: 358 402180131
Email: alifar@utu.fi

Abstract

Background: During the coronavirus disease (COVID-19) pandemic, governments issued movement restrictions and placed areas into quarantine to combat the spread of the disease. In addition, individuals were encouraged to adopt personal health measures such as social isolation. Information regarding the disease and recommended avoidance measures were distributed through a variety of channels including social media, news websites, and emails. Previous research suggests that the vast amount of available information can be confusing, potentially resulting in overconcern and information overload.

Objective: This study investigates the impact of online information on the individual-level intention to voluntarily self-isolate during the pandemic. Using the protection-motivation theory as a framework, we propose a model outlining the effects of cyberchondria and information overload on individuals' perceptions and motivations.

Methods: To test the proposed model, we collected data with an online survey (N=225) and analyzed it using partial least square-structural equation modeling. The effects of social media and living situation were tested through multigroup analysis.

Results: Cyberchondria and information overload had a significant impact on individuals' threat and coping perceptions, and through them on self-isolation intention. Among the appraisal constructs, perceived severity (P=.002) and self-efficacy (P=.003) positively impacted self-isolation intention, while response cost (P<.001) affected the intention negatively. Cyberchondria (P=.003) and information overload (P=.003) indirectly affected self-isolation intention through the aforementioned perceptions. Using social media as an information source increased both cyberchondria and information overload. No differences in perceptions were found between people living alone and those living with their families.

Conclusions: During COVID-19, frequent use of social media contributed to information overload and overconcern among individuals. To boost individuals' motivation to adopt preventive measures such as self-isolation, actions should focus on lowering individuals' perceived response costs in addition to informing them about the severity of the situation.

(J Med Internet Res 2020;22(5):e19128) doi:10.2196/19128

KEYWORDS

COVID-19; pandemic; self-isolation; behavior; protection motivation theory; cyberchondria; information overload

Introduction

Background

The coronavirus disease (COVID-19) pandemic is, in many ways, unique. Compared to the previous worldwide pandemic, the Spanish flu [1], the world has changed significantly. Worldwide trade, travelling, global movement, and the rate at which information is being shared over the internet have all increased drastically. Via the internet, people have access to practically an endless stream of information regarding the new emerging pandemic threat, COVID-19. Through social media, people have shared news articles as well as their own experiences about the pandemic situation, allowing instant access to the latest global developments [2].

Although the vast amount of online data can be useful for artificial intelligence and machine learning algorithms, it can

be difficult for individuals to grasp and conceptualize. We envision two main problems that can emerge from excessive internet use during a worldwide pandemic such as COVID-19: (1) cyberchondria, which is defined as obsessive online searching for health-related information, typically about specific symptoms [3]; and (2) information overload, a condition in which one cannot process all the communications and informational inputs, and as a result, the information gathering process is terminated, or the whole process remains ineffective [4]. Both cyberchondria and information overload have been found to weaken human cognitive reasoning [3-5].

In this study, we investigate how these two factors, cyberchondria and information overload, impact an individual's self-isolation intention during the COVID-19 pandemic. We use the protection motivation theory (PMT) [6] to identify intermediate constructs in between cyberchondria and information overload, and self-isolation intention. We theorized the relationships between these constructs based on previous studies (eg, [6,7]) and formulated a research model. To test the model, we used survey data from Finnish participants (N=225) and analyzed the data using partial least squares-structural equation modeling (PLS-SEM). This was followed by a post hoc analysis regarding the impact of using social media as an information source and living alone. After presenting the results, we describe the theoretical and practical implications of our findings, followed by the limitations and future work. In the end, we provide our conclusions.

Human Behavior During Pandemics and the Case of COVID-19

In late 2019, a highly infectious new virus labelled severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) started spreading in Wuhan, China [8]. In early 2020, this viral respiratory disease had spread to most countries, and on March 11, it was declared a worldwide pandemic by the World Health Organization [9]. As of the middle of April 2020, Johns Hopkins University had reported over 2,000,000 confirmed cases of patients with COVID-19 worldwide [10]. This number is estimated to be a lot higher, as several countries are not testing cases with mild symptoms [9]. The pandemic caused governments to take action, issuing limitations on movements and meetings; closing public services, schools, and universities; and cancelling concerts and other cultural events [11,12].

There are a few established medically proven measures individuals can take during a pandemic to mitigate their chances of contracting the disease: washing hands, avoiding social contact, wearing protective masks, wearing protective gloves, and disinfecting surfaces [13]. These measures were communicated to individuals worldwide through news, social media, and other reports starting from early 2020 when the COVID-19 disease started to become a worldwide issue. In addition to individual-level health protection measures, governments issued orders to avoid large gatherings and placed areas with outbreaks in quarantine [14].

Individual-level behavior during pandemics is a result of both voluntary and government-enforced behavioral change. The benefit of government-enforced measures is that they apply to everyone and have been proven to be effective in crowd control and stopping the spread of diseases [15]. The downsides include negative impacts on the economy and citizens' social well-being. Individuals are more motivated to comply with government-enforced measures and even adapt health measures themselves, if they understand the necessity and reasoning behind the actions [16,17]. A lack of clear communication during unusual, novel, and potentially lethal pandemic situations can lead to uncertainty and even panic among citizens [18]. Accordingly, during pandemics, intervention strategies and information bulletins aiming to propagate health information and knowledge are often used [19].

Theoretical Foundation

One of the most used theories to explain how individuals adopt the promoted health measures, such as self-isolation, is the PMT [6]. The PMT explains individual-level behavioral responses in health-threatening situations [6]. At its core, the theory looks at motivational reasons for adopting protective measures and divides the causes into threat appraisal and coping appraisal. In the context of worldwide pandemics, threat appraisal refers to the individual's perception of the seriousness of the situation, as well as how vulnerable they see themselves and their friends to be in the situation [7,20]. On the other hand, coping appraisal refers to the individual's evaluation of how well they can manage in the given situation. Thus, coping appraisal can be further divided into response costs, self-efficacy, and response efficacy [7].

Previous selected work where the PMT has been used to explain human behavior during pandemics is summarized in Table 1. The literature suggests that there are significant individual differences in the likelihood of adopting health behaviors [20]. Some people feel the need to criticize or neglect suggested health behaviors [21], while others adopt them without complaint. Both threat and coping appraisals have been shown to impact protection motivation [22], with perceived severity being identified as one of the key underlying causes for both appraisals [7,20]. Protection motivation then typically leads to actual behavior [23], but there have been reports to the contrary (eg, [24]).

Despite several studies on pandemic behavior through the lens of the PMT, the existing literature has not exhaustively addressed the impact of internet sources on protection motivation and ultimately behavioral intentions. The role of the internet in pandemic situations is arguably highly complex, as it contains a myriad of information sources and social media platforms through which people can not only acquire knowledge but communicate and share experiences as well. The internet has become the primary source of information for many, but there is a large variance in the preferred online information source. Search engines and social media platforms further complicate the matter with personalized content, which can contribute to some groups of people receiving better and more accurate information regarding the pandemic situation than others. Furthermore, although studies have been done on reactions to several kinds of epidemic situations [20], COVID-19 provides a completely new context, as a pandemic of similar magnitude and impact has not been seen in modern times. To address these research gaps, we looked at cyberchondria and

information overload as internet-specific constructs, and, through the lens of PMT, measured how they affect both threat and coping appraisal and, through them, the intention to self-isolate.

Table 1. The extant literature where PMT has been used to explain behavior during pandemics.

Author(s)	Sample	Disease	Findings
Bish and Michie [20]	Review	Multiple	Older age, being female, being non-white, and education level were associated with increased probability of adopting health behaviors. Personalized intervention strategies were suggested. Perceived threat should be emphasized as well as informing about the effectiveness of protective measures.
McNeill et al [21]	14,312 (tweets)	H1N1	People favored tweets from official sources over unverified sources. However, social media was also used to criticize and question health authorities. Social media played a role in the motivation to adopt health measures.
Miller et al [22]	84	Respiratory infections	Both threat and coping appraisal should be taken into account in interventions, and both can be used to boost protection motivation and cause behavior change.
Sharifirad et al [23]	300	H1N1	Protection motivation lead to adopting preventive behaviors. Perceived severity did not correlate with protection motivation.
Teasdale et al [7]	883	Influenza (general)	Perceived severity influenced both coping and threat appraisal. The coping appraisal was more significant than threat appraisal in determining individuals' actions.
Williams et al [24]	230	Influenza (general)	PMT[a] was useful for explaining intentions to engage in self-isolation behavior, but none of the PMT variables actually lead to adopting these behaviors.

[a]PMT: protection motivation theory.

Research Model and Hypotheses

Cyberchondria and Perceptions

Cyberchondriac behavior is characterized by continuous impulses to go online to find further reading on a concerning health topic. Previous research has identified anxiety and a distaste for ambiguity to be predictors of cyberchondria [25] as well as exposure to too many (contradictory or unclear) information sources [26]. As such, cyberchondria can be regarded to be a product of the internet, as online sources provide a myriad of information on practically any given topic. Accordingly, at times of considerable uncertainty such as the COVID-19 pandemic, an increased amount of cases of cyberchondria can be expected to emerge.

Because the syndrome is fueled by concern for specific health issues or symptoms, it can be expected to increase the perceived severity of the given situation. Online searches for health information lead to finding more information on the topic, and humans tend to look at the worst and scariest cases first. As such, a person suffering from cyberchondria who keeps searching for more information may also experience an increased sense of vulnerability. Thus, we postulated the following hypotheses:

- Hypothesis (H)1: Cyberchondria increases perceived severity.
- H2: Cyberchondria increases perceived vulnerability.

The Impact of Information Overload

Due to the abundance of online and offline information regarding COVID-19, people do not have time to read and understand all available knowledge. When the amount of information crosses one's processing capacities, information overload occurs [27]. The cognitive load theory postulates that the natural human reaction in such situations is to take a step away from the source of the information overload and retreat to a safer ground [28]. This process has successfully been used to explain a wide variety of phenomena, such as student retention in online courses [29].

The human memory is divided into two parts: long-term and working memory. New incoming information is processed by the working memory and kept as schemas in the long-term memory. The working memory can hold only a limited amount of information at a time [28]. The time it takes to process new information is dependent on existing knowledge structures and the degree that conceptual change is required to align the latest information with the existing knowledge [30]. Prior literature has reported that information overload may create fatigue [31] and reduce people's self-regulation ability [32]. Therefore, we expect that information overload may reduce self-efficacy as well as response efficacy. It may also increase the required response costs to make a particular decision because of the feeling of uncertainty resulting from not being able to process all available information. Accordingly, we formulated the following three hypotheses:

- H3: Information overload negatively influences self-efficacy.
- H4: Information overload negatively influences response efficacy.
- H5: Information overload increases perceived response cost.

Factors Affecting Self-Isolation Intention

The PMT postulates that there are two aspects eventually contributing to the motivation to adopt health measures such as self-isolation: threat appraisal and coping appraisal. Building off the work of previous scholars using PMT to investigate behavior during pandemics (eg, [7]), we further divided threat appraisal into perceived severity and perceived vulnerability. Perceived severity has been shown to be perhaps the most

crucial factor leading to protection motivation [20]. Thus, it should positively correlate with the self-isolation intention. Perceived vulnerability has similarly been found to increase taken health measures [33]. Accordingly, we postulated the following hypotheses:

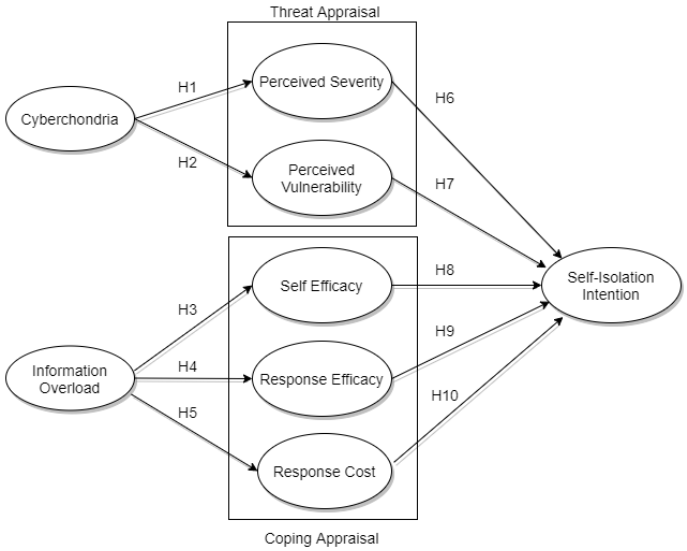
- H6: Perceived severity increases self-isolation intention.
- H7: Perceived vulnerability increases self-isolation intention.

The other part of the PMT, coping appraisal, can be divided into self-efficacy, response efficacy, and response cost [7]. Self-efficacy refers to the individual's beliefs in their capabilities to influence a situation. It also refers to the behavioral skills of a person [34]. On the other hand, response efficacy refers to the perception of one's capability of being able to respond to the situation. For example, in the case of COVID-19, response efficacy consists of the ability to self-isolate at one's own will. Response cost is the individual's evaluation of the negative impact of specific responses. We focused primarily on the self-isolation intention, and thus, the response costs refer to what follows from self-isolation. This could mean in practice losing the opportunity to meet friends or go fishing, or even losing a job and, consequently, income. As the response costs are negative, they should have a strong negative influence on self-isolation intention. Accordingly, we postulated our final three hypotheses.

- H8: Self-efficacy positively influences the self-isolation intention.
- H9: Response efficacy negatively influences the self-isolation intention.
- H10: Response cost negatively influences the self-isolation intention.

The proposed research model with the theorized hypotheses is displayed in Figure 1. The model, which is based on the PMT theory, has two root constructs connected to using the internet for information searches, cyberchondria and information overload. Cyberchondria is linked to two threat appraisal constructs, while information overload is linked to three coping appraisal constructs. All constructs from both threat and coping appraisal are linked to self-isolation intention, which is the sole dependent variable in the model and was selected to present health measures taken during the COVID-19 pandemic. Altogether the model has eight constructs and ten hypotheses.

Figure 1. Research model explaining the relationship of cyberchondria, information overload, and perceptions and intention. H: hypothesis.



Methods

Data Collection

We designed a survey by adapting validated scales from previous literature to measure the constructs. After the survey was drafted, 11 participants were asked to act as a test group to give feedback, ensuring the survey was understandable. At the beginning of the survey, the goal of the study as well as data handling procedures were clearly explained to the participants in a concise manner. Research permission was also formally asked of all participants.

An online survey tool Webropol was used for distributing the survey. The survey link was sent through email lists to students, faculty, and employees at a university in Finland. The survey received 225 responses during the time it was available from March 19-30, 2020. As all survey questions were mandatory, there were no cases of missing data. The responses were screened by the authors to ensure all responses seemed honest. More specifically, we checked if a response was random or intentionally wrong. However, no such cases were reported. Thus, all responses were deemed valid and were included in the analysis. The demographic information of the participants is summarized in Table 2.

Table 2. Participant's (N=225) demographic and background information.

Factors	Distribution, n (%)
Gender	
Female	147 (65.3)
Male	73 (32.4)
Prefer not to tell/nonbinary	5 (2.2)
Age (years)	
≤25	89 (39.5)
26-34	73 (32.4)
35-44	34 (15.1)
≥45	29 (12.9)
Position in university	
Student	148 (65.8)
Faculty	68 (30.2)
Other staff	9 (4.0)
Living situation	
Living alone	122 (54.2)
Living with family/children	103 (45.8)
Source of COVID-19[a] information	
Social media	119 (52.9)
Other channels	106 (47.1)

[a]COVID-19: coronavirus disease.

Measures

Multi-item scales were used to measure cyberchondria, information overload, threat and coping appraisal constructs, and the dependent variable self-isolation intention. All the constructs were measured using a 5-point scale (1=strongly disagree and 5=strongly agree). A total of 25 items were used to measure eight constructs initially. Constructs involved in the study are briefly described next, and item descriptions can be seen in Table 3.

Table 3. Constructs, items, and reliability and validity assessments.

Construct, item	Loading	VIF[a]
Information overload [31] (CR[b]: 0.86; AVE[c]: 0.67)		
"I am often distracted by the excessive amount of information on multiple channels/sources about COVID-19[d]"	0.77	1.453
"I find that I am overwhelmed by the amount of information that I process on a daily basis from multiple channels/sources about COVID-19"	0.85	1.794
"I receive too much information regarding the COVID-19 pandemic to form a coherent picture of what is happening"	0.82	1.481
Cyberchondria [35] (CR: 0.82; AVE: 0.61)		
"After reading information about COVID-19 online, I feel confused"	__[e]	__[e]
"I feel frightened after reading information about COVID-19 online"	0.79	1.381
"I feel frustrated after reading information about COVID-19 online"	0.78	1.478
"Once I start reading information about COVID-19 online, it is hard for me to stop"	0.78	1.265
Perceived severity [36] (CR: 0.70; AVE: 0.52)		
"The negative impact of Coronavirus (COVID-19) is very high"	0.70	1.002
"Coronavirus (COVID-19) can be life-threatening"	__[e]	__[e]
"The Coronavirus (COVID-19) is a serious threat for someone like me"	0.73	1.002
Perceived vulnerability [36] (CR: 0.81; AVE: 0.60)		
"I am vulnerable to contracting Coronavirus (COVID-19) in given circumstances"	0.71	1.367
"I don't think I am likely to get the Coronavirus (COVID-19)"[f]	0.86	1.296.
"I am at risk of catching the Coronavirus (COVID-19)"	0.74	1.567
Self-efficacy [36] (CR: 0.84; AVE:0.64)		
"I am able to take avoidant measures if I want to"	0.79	1.254
"Taking avoidant measures is difficult for me"[f]	0.84	1.652
"Avoidant measures are easy to take"	0.78	1.617
Response efficacy [37] (CR=0.89; AVE=0.80)		
"The avoidant measures are a good way of reducing the risk of contracting Coronavirus (COVID-19)"	0.90	1.614
"The avoidant measures reduce my chance of catching the Coronavirus (COVID-19)"	0.89	1.614
Response cost [37] (CR=0.78; AVE=0.54)		
"The benefits of taking avoidant measures outweigh the costs"[f]	0.72	1.146
"I am discouraged from taking avoidant measures as they would impact my work"	0.75	1.202
"I am discouraged from taking avoidant measures because they feel silly"	0.73	1.215
Self-isolation intention [38] (CR=0.83; AVE=0.55)		
"Deliberately cancel or postpone a social event, such as meeting with friends, eating out, or going to a sports event"	0.77	1.499
"Reduce using public transport"	0.70	1.340
"Avoid going to shops"	0.72	1.455
"Stay at home and study/work remotely"	0.77	1.312

[a]VIF: variance inflation factor.

[b]CR: composite reliability.

[c]AVE: average variance explained.

[d]COVID-19: coronavirus disease.

[e]Items removed due to lower loadings (<0.7).

[f]Items reverse coded for the analysis.

Independent Variables

Cyberchondria was measured using 4 items adapted from [35], whereas 3 items for information overload were adapted from [31].

Threat appraisal was measured using two constructs: perceived severity and perceived vulnerability. Both constructs were measured with the help of 3 items, all adapted from [36].

Coping appraisal was measured in terms of self-efficacy, response efficacy, and response cost. Self-efficacy and response cost were measured with 3 items each, whereas 2 items were used to measure response efficacy. Items for self-efficacy were adapted from [36], whereas items for both response efficacy and response cost were taken from [37].

Dependent Variable

The dependent variable, self-isolation intention, was measured with the help of 4 items measuring avoidance intention adapted from [38]. Before the items, we used the statement "I intend to…".

Data Analysis

The data was downloaded from the survey platform in .csv format, and the initial analysis was carried out in SPSS version 25 (IBM Corp). After initial screening, the data normality was checked using skewness and kurtosis. Some items had values greater than the threshold of 0.3 [39], showing the data was not normally distributed [40]. PLS-SEM has been suggested for data analysis in the case of nonnormal data.

Therefore, data were analyzed using PLS-SEM in SmartPLS 3.2 (SmartPLS GmbH) [41]. In this technique, data is analyzed in two steps. First, the measurement model is tested, ensuring reliability and validity of the constructs involved in the study. Second, an assessment of structural models is carried out, testing relationships between the constructs. As a post hoc analysis, we ran partial least squares-multigroup analysis (PLS-MGA [42] to test differences in the model due to differences in information sources and living arrangements. The post hoc analysis was supported by *t* tests (two-tailed) conducted in SPSS.

Results

Measurement Model Results

As mentioned, the first step in PLS-SEM analysis is to test the reliability and validity of the constructs. Reliability is assessed with internal consistency and items reliability, whereas validity consists of the convergent and discriminant validity. Internal consistency has been traditionally measured using Cronbach alpha; however, composite reliability (CR) has been recommended as a suitable measure of reliability in PLS [42]. Therefore, we considered CR for assessing internal consistency. Item reliability was assessed from the item loadings. Convergent validity was ascertained from the average variance explained (AVE), and discriminant validity was assessed with the Fornell-Larcker criterion [43]. For this purpose, we followed the accepted thresholds recommended by previous studies [41,44].

CR for all the constructs was above the threshold of 0.7 [41]. AVE for all the constructs was above 0.5. Two items, one from cyberchondria and one from perceived severity, were dropped due to item loadings below 0.7 (for details, see Table 3). In the final model, 23 items were used to measure eight constructs. The discriminant validity results based on the Fornell-Larcker criterion are shown in Table 4.

In addition, we also examined the variance inflation factor (VIF) to assess the multicollinearity. The highest VIF was 1.794 (Table 3), which was well below the threshold of 5 [45]. Thus, there was no multicollinearity issue in our data. With these assessments, we concluded that our data had a significant level of convergent and discriminant validity.

Table 4. Discriminant validity using Fornell-Larcker criterion.

Constructs	Self-isolation intention	Cyberchondria	Information overloading	Perceived severity	Perceived vulnerability	Response cost	Response efficacy	Self-efficacy
Self-isolation intention	0.745	—[a]	—	—	—	—	—	—
Cyberchondria	0.210	0.785	—	—	—	—	—	—
Information overloading	–0.02	0.591	0.817	—	—	—	—	—
Perceived severity	0.257	0.396	0.073	0.712	—	—	—	—
Perceived vulnerability	0.062	0.180	0.011	0.200	0.776	—	—	—
Response cost	–0.51	–0.08	0.185	–0.12	–0.01	0.738	—	—
Response efficacy	0.349	0.104	–0.04	0.115	–0.02	–0.49	0.899	—
Self-efficacy	0.396	–0.04	–0.03	–0.04	–0.18	–0.52	0.373	0.800

[a]Not available.

Structural Model Results

Next, we evaluated the proposed research model (see Figure 2). Complete bootstrapping with 5000 subsamples was run for the significance testing. Among the proposed relationships, 7 relationships turned significant. The model explained 34% variance in self-isolation intention.

Structural model statistics (with effect size f^2) for the research model are given in Table 5. An effect size of 0.02 is considered low, 0.15 is medium, and 0.5 is large [41].

We found that both cyberchondria and information overload indirectly impacted self-isolation intention. Cyberchondria had a significant positive effect (b=0.07, t=2.929, P=.003), whereas information overload had a negative effect (b=–0.10, t=3.006, P=.003). Cyberchondria significantly impacts self-isolation intention through perceived severity, whereas information overload has an impact on it through self-efficacy and response cost.

Following the structural model analysis, we conducted several post hoc analyses to understand people's self-isolation intentions. The results are described in the next subsections.

Figure 2. Structural model results. Significant paths are shown with solid lines, with standardized path coefficients (*P<.05, **P<.01), whereas dotted lines show insignificant relationships.

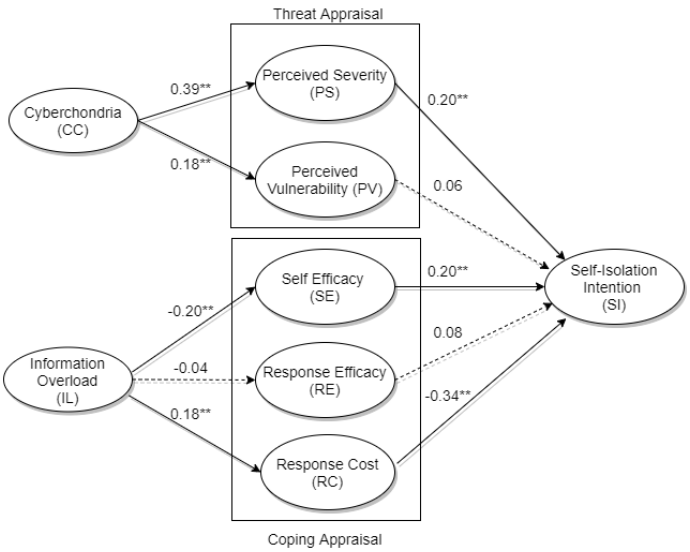


Table 5. Structural model statistics.

Hypothesis	Relationship	β	*t* test	*P* value	f^2
H[a]1 [b]	*CC* [c] *and PS* [d]	*.39*	*7.100*	*<.001*	*0.186*
H2	*CC and PV* [e]	*.19*	*2.639*	*.008*	*0.033*
H3	*IO* [f] *and SE* [g]	*–.20*	*3.218*	*.001*	*0.044*
H4	IO and RE[h]	–.04	0.673	.50	0.002
H5	*IO and RC* [i]	*.18*	*2.828*	*.005*	*0.035*
H6	*PS and SI* [j]	*.20*	*3.139*	*.002*	*0.057*
H7	PV and SI	.05	0.662	.51	0.005
H8	*SE and SI*	*.20*	*2.998*	*.003*	*0.042*
H9	RE and SI	.07	1.067	.24	0.007
H10	*RC and SI*	*–0.34*	*4.374*	*<.001*	*0.112*

[a]H: hypothesis.

[b]Significant relationships are shown in italics.

[c]CC: cyberchondria.

[d]PS: perceived severity.

[e]PV: perceived vulnerability.

[f]IO: Information overload.

[g]SE: self-efficacy.

[h]RE: response efficacy.

[i]RC: response cost.

[j]SI: self-isolation intention.

Effect of Information Source: Post Hoc Analysis

As social media is one of the major sources of information regarding the COVID-19 pandemic, we investigated if the mean values of our constructs, as well as the path coefficients, varied between the respondents who reported social media as the primary source of information and other sources (see Table 6). We observed no significant differences in the beliefs related to threat appraisal and coping appraisal of the respondents who used social media as a primary information source for COVID-19 and those who did not use social media. Similarly, no difference in self-isolation intention was found between the aforementioned groups. The level of cyberchondria and information overload was higher among respondents who used social media as a source to learn about COVID-19 in comparison to the respondents who reported using other channels.

To further see if our model (Figure 1) differs for the respondents who used social media as a source of information and those who used other channels for accumulating COVID-19 related knowledge, we ran PLS-MGA. For PLS-MGA, a *P* value of 0.05 or lower, or 0.95 or higher shows significant path differences in the groups [42].

The R^2 for the self-isolation intention in the social media as an information source group was 0.40, whereas for the other channels groups, it was 0.39. As shown in Table 7, the result of PLS-MGA showed no significant differences between the two groups in most of the relationships. There were only two paths where the difference was significant. First, the effect of self-efficacy on self-isolation intention (H8) was stronger in the group that used social media as an information source compared to the other group. Second, response cost had a stronger effect on self-isolation intention (H10) in the group that used social media as an information source compared to the other group.

Table 6. Difference in beliefs and intention of respondents who use social media and those who use other channels to get information on the coronavirus disease.

Constructs	Social media (n=119), mean (SD)	Other channels (n=106), mean (SD)	*t* test (*df*)	*P* value
Cyberchondria	*2.96 (0.77)*[a]	*2.51 (0.80)*	*4.246 (223)*	*<.001*
Information overloading	*3.15 (0.85)*	*2.70 (0.94)*	*3.779 (223)*	*<.001*
Perceived severity	3.47 (0.63)	3.60 (0.67)	–1.485 (223)	.14
Perceived vulnerability	3.31 (0.84)	3.46 (0.77)	–1.353 (223)	.18
Self-efficacy	3.90 (0.80)	4.09 (0.58)	–1.840 (223)	.07
Response efficacy	4.40 (0.58)	4.45 (0.57)	–0.643 (223)	.52
Response cost	1.79 (0.69)	1.74 (0.59)	0.580 (223)	.56
Self-isolation intention	4.31 (0.60)	4.28 (0.61)	0.371 (223)	.71

[a]Significant differences are shown in italics.

Table 7. Partial least squares-multigroup analysis results for the effect of source of information (as moderator).

Hypothesis	Relationship	Social media (n=119)		Other channels (n=106)		Path difference	*P* value (social media vs other channels)
		t test	*P* value	*t* test	*P* value		
H[a]1	CC[b] and PS[c]	6.707	<.001	3.854	.01	0.111	.35
H2	CC and PV[d]	1.061	.29	2.281	.02	–0.149	.34
H3	IO[e] and SE[f]	0.679	.50	3.734	.01	0.24	.12
H4	IO and RE[g]	0.035	.97	0.839	.40	0.08	.64
H5	IO and RC[h]	1.19	.23	2.616	.009	–0.099	.58
H6	PS and SI[i]	1.561	.12	2.888	.004	–0.131	.27
H7	PV and SI	0.683	.495	1.755	.08	–0.262	.12
H8[j]	*SE and SI*	*3.783*	*<.001*	*0.365*	*.72*	*.29*	*.03*
H9	RE and SI	0.429	.67	1.644	.10	–0.129	.34
H10	*RC and SI*	*2.843*	*.004*	*2.766*	*.006*	*0.002*	*.99*

[a]H: hypothesis.
[b]CC: cyberchondria.
[c]PS: perceived severity.
[d]PV: perceived vulnerability.
[e]IO: Information overload.
[f]SE: self-efficacy.
[g]RE: response efficacy.
[h]RC: response cost.
[i]SI: self-isolation intention.
[j]Significant relationships are shown in italics.

Effect of Living Alone: Post Hoc Analysis

Individuals living in the same household with other people cannot generally avoid contracting diseases to one another. Large households where multiple persons reside can be seen to be at an increased risk of contracting the virus compared to single-person households. Accordingly, we also investigated if the decision making varies between the respondents who live alone compared to those living with others (see Table 8). Interestingly, we did not detect any differences in any of the constructs between the two groups.

In Table 9, we see that 33% of the variance in self-isolation intention was explained by the group living alone, whereas 37% of the variance was explained by living with others. The only path with a significant difference was between information overload and response efficacy, whereas information overload significantly affected response efficacy (negatively) among respondents who lived with others.

Table 8. Difference in beliefs and intention of respondents who live alone and who live with other people.

Constructs	Live alone (n=122), mean (SD)	With others (n=109), mean (SD)	*t* test (*df*)	*P* value
Cyberchondria	2.66 (0.84)	2.86 (0.78)	–1.911 (223)	.06
Information overloading	2.87 (0.93)	3.03 (0.91)	–1.321 (223)	.19
Perceived severity	3.53 (0.68)	3.54 (0.63)	–0.172 (223)	.86
Perceived vulnerability	3.43 (0.81)	3.33 (0.82)	0.918 (223)	.36
Self-efficacy	3.93 (0.73)	4.07 (0.77)	–1.416 (223)	.16
Response efficacy	4.42 (0.56)	4.44 (0.60)	–0.252 (223)	.80
Response cost	1.83 (0.65)	1.71 (0.64)	1.345 (223)	.18
Self-isolation intention	4.23 (0.61)	4.38 (0.60)	–1.863 (223)	.06

Table 9. Partial least squares-multigroup analysis results for the effect of living alone vs with other people (as moderator).

Hypothesis	Relationship	Live alone (n=122)		With others (n=109)		Path difference	*P* value (social media vs other channels)
		t test	*P* value	*t* test	*P* value		
H[a]1	CC[b] and PS[c]	5.365	<.001	2.382	.02	0.107	.47
H2	CC and PV[d]	4.011	<.001	0.441	.66	0.245	.12
H3	IO[e] and SE[f]	2.621	.009	2.228	.03	–0.041	.76
H4	IO and RE[g]	0.65	.52	2.379	.02	0.276	.049
H5	IO and RC[h]	2.426	.02	1.756	.08	0.012	.93
H6	PS and SI[i]	2.13	.03	1.958	.05	–0.029	.85
H7	PV and SI	1.239	.22	0.289	.77	0.136	.34
H8	SE and SI	1.442	.15	2.163	.03	–0.15	.31
H9	RE and SI	1.211	.23	0.491	.62	0.059	.67
H10	RC and SI	3.735	.01	1.826	.07	–0.071	.69

[a]H: hypothesis.

[b]CC: cyberchondria.

[c]PS: perceived severity.

[d]PV: perceived vulnerability.

[e]IO: Information overload.

[f]SE: self-efficacy.

[g]RE: response efficacy.

[h]RC: response cost.

[i]SI: self-isolation intention.

Discussion

Principal Results

Self-isolation intention was predicted through perceived severity (threat appraisal), as well as through self-efficacy (coping appraisal) and response cost (coping appraisal). Unlike we hypothesized, perceived vulnerability and response efficacy did not correlate with the self-isolation intention. When looking at how internet use connects to threat and coping appraisal, we noticed that cyberchondria positively affected perceived severity, whereas information overload had both negative and positive effects on coping appraisal: negative for self-efficacy and positive for response cost. Lastly, both cyberchondria and information overload had an impact on self-isolation intention via the observed intermediate constructs.

The negative impact of information overload on self-efficacy can be explained by the fact that information overload does not allow an accurate understanding of the situation at hand, and uncertainty lowers self-efficacy [46]. The positive influence of information overload on response cost can, in turn, be understood by how the uncertainty that follows from information overload makes it difficult to perceive the situation objectively. As humans have a tendency of assuming the situation to be slightly worse than what it is in reality [47], this leads to an increased perceived response cost.

We also conducted post hoc analyses to study the impacts of using social media as an information source for COVID-19 and

the impact of living alone compared to living with other people. We noticed that social media users experienced a higher level of cyberchondria and information overload; however, this difference did not have a significant effect on our structural model. The above finding suggests that our model is equally suitable for explaining the impact of cyberchondria and information overload on intentions to self-isolate regardless of the information sources people use. Finally, we did not find any significant differences between people living alone compared to people living with others.

Implications of Findings

We report three theoretical contributions from our results. First, we found that information overload distorts people's belief system, particularly coping appraisal during a pandemic. In particular, we found that information overload negatively affected self-efficacy and positively influenced response costs. Therefore, we contribute to the prior literature discussing the outcomes of information overload [31,32] by describing how it interferes with responses during a pandemic.

Second, we observed that cyberchondria influenced the perceptions of perceived severity and perceived vulnerability. We concluded that cyberchondria affects people's threat appraisal during pandemics such as with COVID-19. With this finding, we contribute to the literature on cyberchondria [35] by showing that it plays a significant role in motivating people to adopt recommended health measures. Although cyberchondria is generally regarded to be negative, in the case of COVID-19, it might have helped people understand the actual severity of the situation. However, it also follows from our findings that although people with cyberchondria may be early adopters of self-isolation behavior, they can, in the long run, start to suffer from stress and anxiety due to constantly seeing news and reports highlighting the severity of the situation.

Third, we extend the PMT literature on pandemics (eg, [20,21]) by employing information overload and cyberchondria as predictors of threat and coping appraisal. Prior PMT literature suggests knowledge as an antecedent of threat and coping appraisal [7]. By contrast, our study shows how negative consequences of information (cyberchondria and information overload) shape the threat and coping appraisals. We further proved in our post hoc analysis that using social media as an information source increases both cyberchondria and information overload, which may be explained by the fact that social media news is more subject to individual perceptions and lacks the objective and rigorous approach to information reporting that journalists have.

The findings also have practical implications for health behavior change system designers as well as governments, journalists, and other parties interested in impacting individual-level health behavior. The importance of internet sources and their impact on both threat and coping appraisals must be accounted for when attempting to understand human behavior during pandemics. Interventions targeted to increase an individual's perceived severity to get them to act may unintentionally increase the cyberchondria of those who already perceive the situation to be grave, causing increased strain on people in an already unusual and stressful situation. Supplementing findings from previous studies (eg, [20]), we suggest personalized intervention strategies, where individuals suffering from cyberchondria are given reassuring and hopeful messages, and those with no intention to adopt health measures are targeted with communication that aims to increase their perceived severity of the situation.

We noticed that social media users experience greater levels of cyberchondria and information overload compared to others. The responsibility of online platforms and search engines should, thus, be brought to discussion. Social media sites and search engine developers could take measures to ensure they display clear and comprehensible information to people to avoid the negative consequences of information overload and cyberchondria while still communicating to people the severity of the pandemic and recommended health measures. However, we noted that the content in social media is at large dictated by the people who are using it. Hence, educating people on responsible and healthy social media use could help alleviate the observed negative consequences.

Limitations and Future Work

Our work also has limitations that need to be taken into account. First, the collected data was cross-sectional and hence did not account for any change over time. For example, it is possible that information overload was experienced more heavily at the beginning of the pandemic when the novelty and uncertainty of the situation were greater. Second, participants were selected from a geographically and socially limited area. During the data collection period, Finland had fewer than 1000 confirmed cases of COVID-19, and the country also has a relatively low population density. For increased reliability, our findings could be supplemented from data collected from other countries, especially those that have been hit hard by the pandemic. Third, we chose cyberchondria and information overload as examples of internet-fueled concepts; however, other constructs could have also been used in their place or in addition to them, such as the preference of trusted sources over social media and the impact of hearing experiences about the pandemic directly from people. Finally, in our post hoc analysis, we showed that social media users experienced significantly more cyberchondria and information overload compared to others, but the content on social media was not specified. Thus, future research could look into the specific types of social media behavior and content that contribute to the observed increase in cyberchondria and information overload.

Our findings invite further research to investigate the impacts of internet-related information exposure on health behavior intentions during pandemics. The significant influence of cyberchondria on perceived severity, which further leads to the self-isolation intention, also raises ethical concerns, as a seemingly negative phenomenon (cyberchondria) may be used to motivate people to adopt recommended health measures, in this case, self-isolation. Furthermore, it is possible that continuous behavior interventions aiming to get people to self-isolate may unintentionally fuel cyberchondria for those already worried. Accordingly, behavioral intervention campaigns aiming to get the entirety of the population to self-isolate

voluntarily may, after a certain threshold, begin to cause more harm than good.

Supplementing previous studies on the topic, the importance of accurate, precise, and reliable information was highlighted by our findings. The government-level quarantine and movement restrictions that were placed during the COVID-19 pandemic presented a novel research problem, which has not yet been studied, and that is how online information sources impact the adoption of recommended health behaviors during pandemics. In the case of Finland, during the time of the data collection, all participants had been ordered to work from home. This most likely translates to increased time spent on computers and online, which together with the novel and uncertain COVID-19 situation may likely have increased the amount of information read online related to the pandemic. As such, the quarantine measures may be effective at stopping the spread of the pandemic, but they may boost unhealthy internet behavior, most acutely, cyberchondria.

Conclusions

The purpose of this study was to examine the effects of information overload and cyberchondria, two constructs measuring consequences of online information, on self-isolation intention during the COVID-19 pandemic. To understand this effect, we used the PMT framework. Accordingly, we constructed a model where we measured the impact of information overload and cyberchondria on the coping and threat appraisal construct of PMT, and their subsequent relationship with the self-isolation intention. The results revealed perceived severity and self-efficacy to positively influence self-isolation intention, while response cost had a negative effect on it. Both cyberchondria and information overload were found to be higher among those who used social media as an information source. It follows from our findings that, although cyberchondria and information overload are generally regarded to be negative, during the COVID-19 pandemic they contributed to the adoption of recommended health behavior (self-isolation). The finding that using social media as an information source increases both cyberchondria and information overload invites further research into the impact of social media on human behavior during pandemics. Finally, our findings suggest that intervention strategies motivating people to adopt health measures should focus not only on stressing the severity of the situation but also on reducing information overload via the clear structuring and communication of reliable health information. This can help mitigate cyberchondria, which we showed may arise as a side effect of a severe worldwide pandemic such as COVID-19.

Conflicts of Interest

None declared.

References

1. Gibbs M, Armstrong J, Gibbs A. Recombination in the hemagglutinin gene of the 1918 "Spanish flu". Science 2001 Sep 07;293(5536):1842-1845. [doi: 10.1126/science.1061662] [Medline: 11546876]
2. Sharma M, Yadav K, Yadav N, Ferdinand KC. Zika virus pandemic-analysis of Facebook as a social media health information platform. Am J Infect Control 2017 Mar 01;45(3):301-302. [doi: 10.1016/j.ajic.2016.08.022] [Medline: 27776823]
3. Fergus TA, Russell LH. Does cyberchondria overlap with health anxiety and obsessive-compulsive symptoms? An examination of latent structure and scale interrelations. J Anxiety Disord 2016 Mar;38:88-94. [doi: 10.1016/j.janxdis.2016.01.009] [Medline: 26851749]
4. Breaudoin CE. Explaining the relationship between internet use and interpersonal trust: taking into account motivation and information overload. J Comput Commun 2008 Apr 01;13(3):550-568. [doi: 10.1111/j.1083-6101.2008.00410.x]
5. Starcevic V, Berle D. Cyberchondria: towards a better understanding of excessive health-related Internet use. Expert Rev Neurother 2013 Feb;13(2):205-213. [doi: 10.1586/ern.12.162] [Medline: 23368807]
6. Rogers RW, Prentice-Dunn S. Protection motivation theory. In: Gochman DS, editor. Handbook of Health Behavior Research 1: Personal and Social Determinants. New York, NY: Plenum Press; 1997:113-132.
7. Teasdale E, Yardley L, Schlotz W, Michie S. The importance of coping appraisal in behavioural responses to pandemic flu. Br J Health Psychol 2012 Feb;17(1):44-59. [doi: 10.1111/j.2044-8287.2011.02017.x] [Medline: 22233104]
8. Wölfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Müller MA, et al. Virological assessment of hospitalized patients with COVID-2019. Nature 2020 Apr 01:e. [doi: 10.1038/s41586-020-2196-x] [Medline: 32235945]
9. World Health Organization. 2020. Coronavirus disease (COVID-19) Pandemic URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [accessed 2020-04-02]
10. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. The Lancet Infectious Diseases 2020 May;20(5):533-534. [doi: 10.1016/s1473-3099(20)30120-1]
11. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? The Lancet 2020 Mar;395(10228):931-934. [doi: 10.1016/s0140-6736(20)30567-5]
12. 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]

13. Chen H, Guo J, Wang C, Luo F, Yu X, Zhang W, et al. Clinical characteristics and intrauterine vertical transmission potential of COVID-19 infection in nine pregnant women: a retrospective review of medical records. Lancet 2020 Mar 07;395(10226):809-815 [FREE Full text] [doi: 10.1016/S0140-6736(20)30360-3] [Medline: 32151335]
14. Remuzzi A, Remuzzi G. COVID-19 and Italy: what next? The Lancet 2020 Apr;395(10231):1225-1228. [doi: 10.1016/s0140-6736(20)30627-9]
15. Maier B, Brockmann D. Effective containment explains sub-exponential growth in confirmed cases of recent COVID-19 outbreak in Mainland China. arXiv preprint 2020:9 (forthcoming) [FREE Full text] [doi: 10.1101/2020.02.18.20024414]
16. Cowper A. Covid-19: are we getting the communications right? BMJ 2020 Mar 06;368:2. [doi: 10.1136/bmj.m919] [Medline: 32144115]
17. Deci EL, Ryan RM. Intrinsic motivation. In: Weiner I, Craighead WE, editors. The Corsini Encyclopedia Of Psychology. New Jersey: John Wiley & Songs; 2010:2.
18. Wang CJ, Ng CY, Brook RH. Response to COVID-19 in Taiwan: big data analytics, new technology, and proactive testing. JAMA 2020 Mar 03:e. [doi: 10.1001/jama.2020.3151] [Medline: 32125371]
19. Liu B, Kim S. How organizations framed the 2009 H1N1 pandemic via social and traditional media: Implications for U.S. health communicators. Public Relations Review 2011 Sep;37(3):233-244. [doi: 10.1016/j.pubrev.2011.03.005]
20. Bish A, Michie S. Demographic and attitudinal determinants of protective behaviours during a pandemic: a review. British Journal of Health Psychology 2010;15(4):797-824. [doi: 10.1348/135910710x485826]
21. McNeill A, Harris PR, Briggs P. Twitter influence on UK vaccination and antiviral uptake during the 2009 H1N1 pandemic. Front Public Health 2016;4:26. [doi: 10.3389/fpubh.2016.00026] [Medline: 26942174]
22. Miller S, Yardley L, Little P, PRIMIT team. Development of an intervention to reduce transmission of respiratory infections and pandemic flu: measuring and predicting hand-washing intentions. Psychol Health Med 2012;17(1):59-81. [doi: 10.1080/13548506.2011.564188] [Medline: 21644184]
23. Sharifirad G, Yarmohammadi P, Sharifabad MA, Rahaei Z. Determination of preventive behaviors for pandemic influenza A/H1N1 based on protection motivation theory among female high school students in Isfahan, Iran. J Educ Health Promot 2014;3:7 [FREE Full text] [doi: 10.4103/2277-9531.127556] [Medline: 24741647]
24. Williams L, Rasmussen S, Kleczkowski A, Maharaj S, Cairns N. Protection motivation theory and social distancing behaviour in response to a simulated infectious disease epidemic. Psychol Health Med 2015;20(7):832-837. [doi: 10.1080/13548506.2015.1028946] [Medline: 25835044]
25. Norr AM, Albanese BJ, Oglesby ME, Allan NP, Schmidt NB. Anxiety sensitivity and intolerance of uncertainty as potential risk factors for cyberchondria. J Affect Disord 2015;174:64-69. [doi: 10.1016/j.jad.2014.11.023] [Medline: 25486275]
26. Loos A. Cyberchondria: too much information for the health anxious patient? J Consum Health Internet 2013 Oct;17(4):439-445. [doi: 10.1080/15398285.2013.833452]
27. Bawden D, Robinson L. The dark side of information: overload, anxiety and other paradoxes and pathologies. Journal of Information Science 2008 Nov 21;35(2):180-191. [doi: 10.1177/0165551508095781]
28. Sweller J. Cognitive load theory. In: Mestre JP, Ross BH, editors. Psychology of Learning and Motivation. California: Elsevier; 2011:37-76.
29. Laato S, Lipponen E, Salmento H, Vilppu H, Murtonen M. Minimizing the number of dropouts in university pedagogy online courses. 2019 Presented at: CSEDU 2019-Proceedings of the 11th International Conference on Computer Supported Education; 2019; Crete, Greece p. 587. [doi: 10.5220/0007686005870596]
30. Cook MP. Visual representations in science education: the influence of prior knowledge and cognitive load theory on instructional design principles. Sci Ed 2006 Nov;90(6):1073-1091. [doi: 10.1002/sce.20164]
31. Whelan E, Islam AN, Brooks S. Applying the SOBC paradigm to explain how social media overload affects academic performance. Comput Educ 2020 Jan;143:103692. [doi: 10.1016/j.compedu.2019.103692]
32. Whelan E, Najmul Islam A, Brooks S. Is boredom proneness related to social media overload and fatigue? A stress–strain–outcome approach. INTR 2020 Feb 24;ahead-of-print(ahead-of-print):19 (forthcoming). [doi: 10.1108/intr-03-2019-0112]
33. Bults M, Beaujean DJ, de Zwart O, Kok G, van Empelen P, van Steenbergen JE, et al. Perceived risk, anxiety, and behavioural responses of the general public during the early phase of the Influenza A (H1N1) pandemic in the Netherlands: results of three consecutive online surveys. BMC Public Health 2011 Jan 03;11:2. [doi: 10.1186/1471-2458-11-2] [Medline: 21199571]
34. Farooq A, Jeske D, Isoaho J. Predicting students' security behavior using information-motivation-behavioral skills model. In: Dhillon G, Karlsson F, Hedstrom K, Zuquete A, editors. ICT Systems Security and Privacy Protection. Lisbon, Portugal: Springer; 2019:238-252.
35. Jokić-Begić N, Mikac U, Čuržik D, Sangster Jokić C. The development and validation of the short cyberchondria scale (SCS). J Psychopathol Behav Assess 2019 May 17;41(4):662-676. [doi: 10.1007/s10862-019-09744-z]
36. Ling M, Kothe EJ, Mullan BA. Predicting intention to receive a seasonal influenza vaccination using Protection Motivation Theory. Soc Sci Med 2019 Jul;233:87-92. [doi: 10.1016/j.socscimed.2019.06.002] [Medline: 31195194]
37. Milne S, Orbell S, Sheeran P. Combining motivational and volitional interventions to promote exercise participation: protection motivation theory and implementation intentions. Br J Health Psychol 2002 May;7(Pt 2):163-184. [doi: 10.1348/135910702169420] [Medline: 14596707]

38. Rubin GJ, Amlôt R, Page L, Wessely S. Public perceptions, anxiety, and behaviour change in relation to the swine flu outbreak: cross sectional telephone survey. BMJ 2009 Jul 02;339:b2651 [FREE Full text] [doi: 10.1136/bmj.b2651] [Medline: 19574308]
39. Kline R. Principles and Practice of Structural Equation Modeling. New York: The Guilford Press; 2016:978-971.
40. Sarstedt M, Ringle C, Smith D, Reams R, Hair J. Partial least squares structural equation modeling (PLS-SEM): a useful tool for family business researchers. J Fam Bus Strateg 2014 Mar;5(1):105-115. [doi: 10.1016/j.jfbs.2014.01.002]
41. Hair J, Hult G, Ringle C, Sarstedt M. A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). California, USA: SAGE publications, Inc; 2017.
42. Henseler J, Ringle C, Sinkovics R. The use of partial least squares path modeling in international marketing. In: Sinkovics R, Ghauri P, editors. New Challenges to International Marketing. Bingley, UK: Emerald Group Publishing Ltd; 2009.
43. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Mark Res 2018 Nov 28;18(1):39-50. [doi: 10.1177/002224378101800104]
44. Henseler J, Ringle CM, Sarstedt M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. J of the Acad Mark Sci 2014 Aug 22;43(1):115-135. [doi: 10.1007/s11747-014-0403-8]
45. Hair Jr JF, Sarstedt M, Hopkins L, Kuppelwieser VG. Partial least squares structural equation modeling (PLS-SEM). Eur Bus Rev 2014 Mar 04;26(2):106-121. [doi: 10.1108/ebr-10-2013-0128]
46. Kim B, Rhee E, Ha G, Yang J, Lee S. Tolerance of uncertainty: links to happenstance, career decision self-efficacy, and career satisfaction. Career Dev Q 2016 May 31;64(2):140-152. [doi: 10.1002/cdq.12047]
47. Thornton A. Optimism vs pessimism: design decisions in the face of process capability uncertainty. J Mech Des 2001;123(3):313-321. [doi: 10.1115/1.1371774]

Abbreviations

AVE: average variance explained
COVID-19: coronavirus disease
CR: composite reliability
H: hypothesis
PLS-MGA: partial least squares-multigroup analysis
PLS-SEM: partial least squares-structural equation modeling
PMT: protection motivation theory
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
VIF: variance inflation factor

Edited by G Eysenbach; submitted 04.04.20; peer-reviewed by H Makkonen, S Talukder; comments to author 17.04.20; revised version received 20.04.20; accepted 21.04.20; published 06.05.20.

Please cite as:
Farooq A, Laato S, Islam AKMN
Impact of Online Information on Self-Isolation Intention During the COVID-19 Pandemic: Cross-Sectional Study
J Med Internet Res 2020;22(5):e19128
URL: http://www.jmir.org/2020/5/e19128/
doi:10.2196/19128
PMID:32330115

Original Paper

COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data

Wasim Ahmed[1], BA, MSc, PhD; Josep Vidal-Alaball[2,3], MD, PhD; Joseph Downing[4], BSc, MSc, PhD; Francesc López Seguí[5,6], MSc

[1]Newcastle University, Newcastle upon Tyne, United Kingdom

[2]Health Promotion in Rural Areas Research Group, Gerència Territorial de la Catalunya Central, Institut Català de la Salut, Sant Fruitós de Bages, Spain

[3]Unitat de Suport a la Recerca de la Catalunya Central, Fundació Institut Universitari per a la Recerca a l'Atenció Primària de Salut Jordi Gol i Gurina, Sant Fruitós de Bages, Spain

[4]London School of Economics, European Institute, London, United Kingdom

[5]TIC Salut Social, Generalitat de Catalunya, Barcelona, Spain

[6]CRES & CEXS, Universitat Pompeu Fabra, Barcelona, Spain

Corresponding Author:
Wasim Ahmed, BA, MSc, PhD
Newcastle University
5 Barrack Rd
Newcastle upon Tyne, NE1 4SE
United Kingdom
Phone: 44 (0) 191 208 150
Email: Wasim.Ahmed@Newcastle.ac.uk

Abstract

Background: Since the beginning of December 2019, the coronavirus disease (COVID-19) has spread rapidly around the world, which has led to increased discussions across online platforms. These conversations have also included various conspiracies shared by social media users. Amongst them, a popular theory has linked 5G to the spread of COVID-19, leading to misinformation and the burning of 5G towers in the United Kingdom. The understanding of the drivers of fake news and quick policies oriented to isolate and rebate misinformation are keys to combating it.

Objective: The aim of this study is to develop an understanding of the drivers of the 5G COVID-19 conspiracy theory and strategies to deal with such misinformation.

Methods: This paper performs a social network analysis and content analysis of Twitter data from a 7-day period (Friday, March 27, 2020, to Saturday, April 4, 2020) in which the #5GCoronavirus hashtag was trending on Twitter in the United Kingdom. Influential users were analyzed through social network graph clusters. The size of the nodes were ranked by their betweenness centrality score, and the graph's vertices were grouped by cluster using the Clauset-Newman-Moore algorithm. The topics and web sources used were also examined.

Results: Social network analysis identified that the two largest network structures consisted of an isolates group and a broadcast group. The analysis also revealed that there was a lack of an authority figure who was actively combating such misinformation. Content analysis revealed that, of 233 sample tweets, 34.8% (n=81) contained views that 5G and COVID-19 were linked, 32.2% (n=75) denounced the conspiracy theory, and 33.0% (n=77) were general tweets not expressing any personal views or opinions. Thus, 65.2% (n=152) of tweets derived from nonconspiracy theory supporters, which suggests that, although the topic attracted high volume, only a handful of users genuinely believed the conspiracy. This paper also shows that fake news websites were the most popular web source shared by users; although, YouTube videos were also shared. The study also identified an account whose sole aim was to spread the conspiracy theory on Twitter.

Conclusions: The combination of quick and targeted interventions oriented to delegitimize the sources of fake information is key to reducing their impact. Those users voicing their views against the conspiracy theory, link baiting, or sharing humorous tweets inadvertently raised the profile of the topic, suggesting that policymakers should insist in the efforts of isolating opinions that are based on fake news. Many social media platforms provide users with the ability to report inappropriate content, which should be used. This study is the first to analyze the 5G conspiracy theory in the context of COVID-19 on Twitter offering practical guidance to health authorities in how, in the context of a pandemic, rumors may be combated in the future.

(J Med Internet Res 2020;22(5):e19458) doi:10.2196/19458

KEYWORDS

COVID-19; coronavirus; twitter; misinformation; fake news; 5G; social network analysis; social media; public health; pandemic

Introduction

The coronavirus strains have been known since 1960 and usually cause up to 15% of common colds in humans each year, mainly in mild forms. Previously two variants of coronavirus have caused severe illnesses: severe acute respiratory syndrome (SARS) in 2002, with severe acute respiratory distress, resulting in 9.6% mortality; and Middle East respiratory syndrome in 2012, with a higher mortality rate of 34.4% [1-3]. The novel coronavirus (SARS coronavirus 2), the seventh coronavirus known to infect humans, is a positive single-stranded RNA virus that probably originated in a seafood market in Wuhan in December 2019 [4,5]. Since then, the coronavirus disease (COVID-19), named by the World Health Organization, has affected more than 2 million people worldwide, killing more than 130,000 of them [6]. The COVID-19 pandemic coincided with the launch and development of the 5G mobile network.

Compared to the current 4G networks, 5G wireless communications provide high data rates (ie, gigabytes per second), have low latency, and increase base station capacity and perceived quality of service [7]. The popularity of this technology arose because of the burst in smart electronic devices and wireless multimedia demand, which created a burden on existing networks. A key benefit of 5G is that some of the current issues with cellular networks such as poor data rates, capacity, quality of service, and latency will be solved [7]. Although there is no scientific proof, the technology is suggested to negatively affect health on certain social media channels [8].

In the first week of January, some social media users pointed to 5G as being the cause of COVID-19 or accelerating its spread. The issue became a trending topic and appeared visible to all users on Twitter within the United Kingdom. Since then, multiple videos and news articles have been shared across social media linking the two together. The conspiracy has been of such a serious nature that, in Birmingham and Merseyside, United Kingdom, 5G masts were torched over concerns associating this technology and the spread of the disease according to the British Broadcasting Corporation [9]. More concerningly, Nightingale hospital in Birmingham, United Kingdom had its phone mast set on fire [10]. This is unwelcome damage especially at a time when hospitals are required to operate with maximum efficiency.

The independent fact-checking website Full Fact noted that the conspiracy was not true and concluded that the theories given to support the 5G claims were flawed [11]. The National Health Service Director, Stephen Powis, noted in a press conference that the 5G infrastructure is vital for the wider general population who are being asked to remain at home. He noted that: “I'm absolutely outraged and disgusted that people would be taking action against the infrastructure we need to tackle this emergency” [10].

The origin of this theory demonstrates the transnational dimension to the new media landscape and the way that fake news and conspiracy theories travel. Previous research has traced the emergence of the conspiracy theory to comments made by a Belgian doctor in January 2020, linking health concerns about 5G to the emergence of the coronavirus [12]. From April 2-6, 2020, it is estimated that at least 20 mobile phone masts were vandalized in the United Kingdom alone [13]. Social media is an important information source for a subset of the population, and previous seminal research has noted the potential of Twitter for providing real time content analysis, allowing public health authorities to rapidly respond to concerns raised by the public [14]. During the unfolding COVID-19 pandemic, recent research has found that platforms such as YouTube have immense reach and can be used to educate the public [15]. Furthermore, recent research has also called for more understanding of public reactions on social media platforms related to COVID-19 [15].

The aim of this study was to analyze the 5G and COVID-19 conspiracy theory. More specifically, the research objectives were to answer the following questions: (1) who is spreading this conspiracy theory on Twitter; (2) what online sources of information are people referring to; (3) do people on Twitter really believe 5G and COVID-19 are linked; and (4) what steps and actions can public health authorities take to mitigate the spread of this conspiracy theory?

Methods

The data set used in this article consists of 6556 Twitter users whose tweets contained the “5Gcoronavirus” keyword or the #5GCoronavirus hashtag, or were replied to or mentioned in these tweets from Friday, March 27, 2020, at 19:44 Coordinated Universal Time (UTC) to Saturday, April 4, 2020, at 10:38 UTC. Users were included in the data set if they sent a tweet during the time the data was retrieved or were mentioned or replied to in these tweets. This specific keyword and hashtag were selected, as this was the most popular and briefly became a trending topic on Twitter within the United Kingdom in early April. The network consists of a total of 10,140 tweets, which are composed of 1938 mentions, 4003 retweets, 759 mentions in retweets, 1110 replies, and 2328 individual tweets. The data was retrieved using NodeXL (Social Media Research Foundation) and the network graph was laid out using the Harel-Koren Fast Multiscale layout algorithm [16]. In interpreting the network graph, the results build upon previous seminal research, which has identified six network shapes and structures that Twitter topics tend to follow [17]. These network shapes can consist of broadcast networks, polarized crowds, brand clusters, tight crowds, community clusters, and support networks. A computer running Microsoft Windows 8 was used to retrieve data in Microsoft Excel 2010 using the professional version of NodeXL (release code: +1.0.1.428+). NodeXL uses Twitter’s search application programming interface (API). URLs were automatically expanded within NodeXL.

A number of techniques were drawn upon. First, the study used the 5Gcoronavirus keyword, which retrieved mentions of both "5Gcoronavirus" and "#5Gcoronavirus." Second, influential users, topics, and web sources were studied, and a social network analysis of the discussion was conducted with NodeXL, a validated methodology used in previous research [18,19], which provided an understanding of the shape of the conversation. The graph's vertices were grouped by cluster using the Clauset-Newman-Moore algorithm. Third, a manual content analysis [20] of Twitter data was conducted by removing a 10.00% sample of individual tweets (n=233/2328). Coding categories were created by exploring the data and the extracted sample was read and coded. In our content analysis, mentions were not examined because they are typically conversations between users, and retweets were excluded to avoid overpopulating the sample with similar messages. Retweets and mentions were only removed for the manual content analysis and all other analysis in the study includes them. Only English-language tweets were coded. The coding was confirmed by another author and any disagreements were discussed and resolved, which led to a 100% agreement.

Individual users have been anonymized in-line with widely cited best practices for research on Twitter [21].

Results

Social Network Analysis

Figure 1 groups Twitter users in social network graph clusters. Each small color dot represents a user and a line between them represents an edge. Groups were formed around this topic based on how frequently users mentioned each other. There is an edge for each "replies-to" relationship in a tweet, an edge for each "mentions" relationship in a tweet, and a self-loop edge for each tweet that is not a "replies-to" or "mentions." The size of the nodes has been ranked by their betweenness centrality score (BCS) [22], which measures the influence of a vertex over the flow of information between all other vertices under the assumption that information flows over the shortest paths among them.

Figure 1. Social network graph of "5Gcoronavirus".

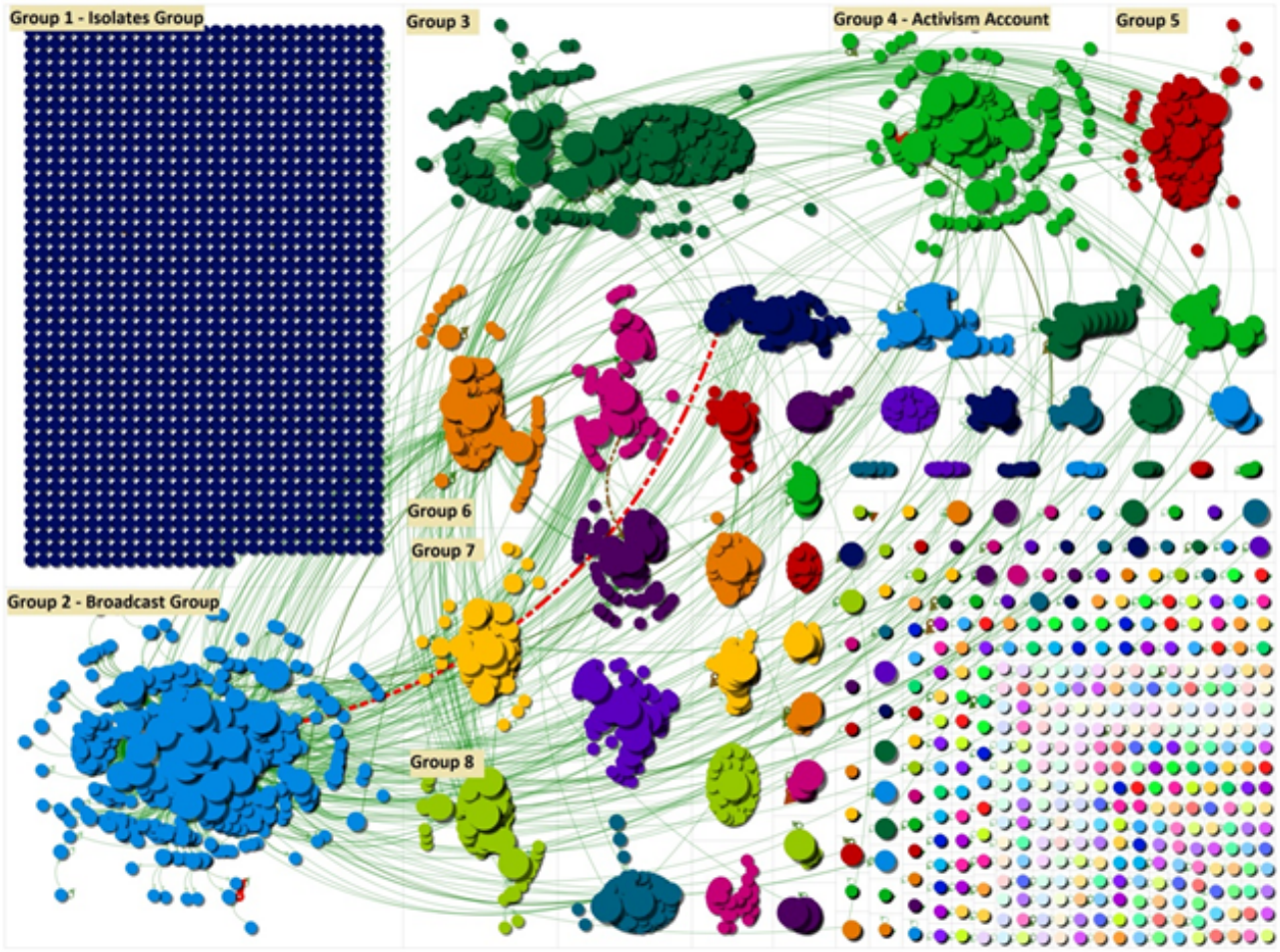


Figure 1 highlights that a number of different groups were formed, but two large groups stand out within the cluster, which are labelled as "*Group 1 - Isolates Group*" and ""*Group 2 - Broadcast Group*." The network shape "*Group 1 - Isolates Group*" displays users who were tweeting without mentioning one another. Isolates groups are a common network structure found in Twitter networks. Large brands, sporting events, and breaking news stories tend to have a large isolates network structure. During a sports event, for example, a large number of users may offer their view or opinion toward a team without mentioning or replying to other users forming an isolates cluster. Group 2 (labeled Group 2 - *Broadcast Group)* contains a number

of Twitter accounts who would tweet that there was a link between 5G and COVID-19, which attracted retweets, giving rise to a broadcast network. Within this group, a number of influential user accounts can also be seen toward the center of the group and a circle of accounts around these. The broadcast network structure is often found in the networks for news accounts and journalists because their tweets are retweeted with high frequency. Celebrities with large followings will also tend to have a broadcast network shape. Group 4 contains the label "*activism account*" because it contained an account with the name "5gcoronavirus19," which was set up to spread the conspiracy theory on Twitter that is further discussed in the next section.

User Analysis

Table 1 has ranked influential users by betweenness centrality and has provided a description of the user account. The rank column orders users by their betweenness centrality score, the account description provides an outline of the type of account, the betweenness centrality column provides the raw score for each user, the follower column lists the number of followers an account had, and the NodeXL group column identifies which group Twitter users belonged to in Figure 1. The follower count is based on the amount of followers the users had during this time period.

Table 1. Influential users ranked by their betweenness centrality score.

Rank	Account description	Betweenness centrality score	Followers, n	Network group in Figure 1
1	Citizen	3,059,934.33	432	7
2	Citizen	3,042,916.47	12	2
3	Citizen	2,926,695.58	546	3
4	Writer	2,655,235.44	1874	2
5	5G and coronavirus dedicated activism account	2,637,433.23	383	4
6	Citizen	2,577,072.58	14	6
7	Citizen	2,354,744.84	175	2
8	Citizen	2,066,430.77	51	2
9	YouTuber	2,003,753.23	130	5
10	Donald Trump	1,380,314.74	75,916,289	4

The majority of influential users tweeting about 5G and COVID-19 consisted of members of the public sharing their views and opinions or news articles and videos supporting their cause. A key feature of the accounts was that they were actively engaged in sharing conspiracy theories; their bios included words such as "uncover" and "truth."

User accounts ranked 1-3 appeared to be citizens who were tweeting during this time. The fourth most influential user was a writer who had over 1874 followers. Interestingly, results show that the fifth most influential account was a dedicated propaganda account (created on January 24, 2012), whose sole purpose was to raise awareness of the link between COVID-19 and 5G, and the account was named "5gcoronavirus19." This account was in the group labeled "*activism account*" in Figure 1. The account creation date appears to be 2012, which suggests that a previously created account was converted, as Twitter allows users to change their user handle and username. The account has since been removed; however, the account bio description was "*5G causes our immune system to lower and we become more susceptible to viruses. Wuhan was the FIRST FULL 5G city! #Coronavirus caused by 5G*." This user was in group 4 in the network graph outlined in Figure 1 and had sent a total of 303 tweets in the 7-day time period studied in this paper. Group 4 contained a total of 408 Twitter accounts. At tenth place, the president of the United States, Donald Trump, appears as an influential user; however, unlike other users in the network, Trump did not directly tweet about the link between COVID-19 and 5G. Trump appears because he is mentioned by other Twitter users related to general policy and discussion surrounding 5G. All other users within the network were actively tweeting during this time period.

The analysis reveals that there was a lack of an authority figure who was actively combating such misinformation. Twitter users that have the highest number of mentions during this time period are shown in Multimedia Appendix 1. The two most mentioned users were members of the public, and the third most mentioned user was the dedicated COVID-19 activism account mentioned previously and ranked the fifth most influential user in Table 1. Multimedia Appendix 2 lists the users that were replied to the most during this time period. The second most mentioned user was again the dedicated coronavirus activism account. This shows how the account linking 5G to COVID-19 also stimulated debate on Twitter and held power over the network because the account was both highly influential and mentioned.

Influential English-Language Websites

In order for the conspiracy theory to spread the public needs information and a reference source. This study has identified which sources were influential during this time. Table 2 highlights the most influential websites relating to this topic during this time period. The most popular web source shared on Twitter during this time was the website known as *InfoWars*, which is a popular conspiracy theory website based in the United States. The article itself linked to several videos in which "top scientists" revealed how 5G could weaken a population's

immune system. The rank column refers to the ranking of each website based on the count column.

It can be seen that the majority of the websites can be argued as being "fake" or "alternative" news websites. The websites and information shared on Twitter can also shed light on the types of sources social media users were drawn toward. Multimedia Appendix 3 shows the most frequently occurring domains within the network. This analysis is different to that conducted in Table 2, as it identifies overall web domains that were most used in tweets rather than specific websites, showing, for instance, that YouTube appears ranked as the second most popular domain.

Table 2. Influential web sources.

Rank	Website title	Source	Tweets, n
1	"WATCH LIVE: CORONAVIRUS HOME SCHOOL SPECIAL & ASK THE EXPERTS! Prestigious doctors & scientists confirm 5G weakens the immune system to all viruses including Covid-19"	InfoWars [23]	38
2	"VIDEO: Former President Of Microsoft Canada, Frank Clegg: 5G Wireless IS NOT SAFE"	RayGuardNJ Electrosmog Protection [24]	31
3	"There's A Connection Between Coronavirus And 5G"	Stillnessinthestorm [25]	20
4	"BREAKING NEWS: Slovenia Stops 5G Due to Health Risks"	5gcrisis website [26]	18
5	"CAN YOU BELIEVE THIS??"	Jeff Censored! YouTube channel. [27]	18

Content Analysis

From the overall data set, a 10.00% sample of tweets (n=233/2328) that did not mention or reply to another user were extracted. Content analysis revealed that, of the 233 sample tweets, 34.8% (n=81) of individual tweets contained views that 5G and COVID-19 were linked, 32.2% (n=75) denounced the conspiracy theory, and 33.0% (n=77) were general tweets not expressing any personal views or opinions. Table 3 below displays the results of this coding alongside examples of tweets.

The focus was to identify the percent of pro- and anticonspiracy themes. Any other tweets would be classified as "general tweets." It was found that 32.2% (n=75/233) of tweets were views against the conspiracy theories that were being shared. They either attacked or ridiculed those sharing such views with humor.

The second category contained tweets that were general in nature and used the "5G" keyword or hashtag in their tweets as highlighted in Table 3. This occurred in 33.0% (n=77/233) of tweets. Users may have used the keywords and hashtags for additional exposure. This theme also contained general news articles related to 5G and COVID-19. This is not surprising, as other Twitter users attempt to "link bait" on Twitter by flooding popular topics with content to obtain more viewers for their own tweets or web links.

The next category consisted of tweets that were clearly expressing views against the conspiracy or were intending to be humorous toward those linking 5G and COVID-19.

The largest category of users, with 34.8% (n=81/233) of the tweets, were engaging with and spreading information that linked COVID-19 and 5G. Anonymized tweet extracts for this theme are provided in Textbox 1.

Thus, 65.2% (n=152/233) of tweets derived from nonconspiracy theory supporters, which suggests that, although the topic attracted high volume, only a handful of users genuinely believed the conspiracy. It is also worth noting that on April 4, 2020, the media began to report that a number of 5G masts had been set on fire [8]. This coincides with the final day that we collected data, and we observed users actively encouraging other users to destroy 5G towers, as highlighted by the final three anonymized tweet extracts in Textbox 1.

Table 3. Content analysis of individual tweets (n=233).

Category	Theme	Example	Tweets, n (%)
1	5G and the coronavirus disease are linked	"*5G Kills! #5Gcoronavirus - they are linked! People don't be blind to the truth!*"	81 (34.8)
2	General tweets not expressing a view or opinion	"*I have a 10AM Skype Chat on Monday, COVID-19 #5Gcoronavirus*"	77 (33.0)
3	Anticonspiracy theories or humor	"*5G is not harming or killing a single person! COVID-19 #5Gcoronavirus*"	75 (32.2)

Textbox 1. Anonymized tweet extracts from category 1.

Tweets
"5G is the one and only Coronavirus! Radiation from it will easily wipe out the world population. Think! Why did China get rid of their 5G towers? This is why they are now free from the Corona."
"5G volumes peaked and infected COVID-19 cases in Italy also peaked, no coincidence!"
"People must open their minds and see the truth that 5G kills!"
"I didn't believe in all of this stuff until I read this article! [URL] Folks, please educate yourselves!"
"Make sure to SMASH THOSE 5G masts up!! #5Gcoronavirus"
"5G Towers are burning [link to video] - now what should we do with the others?"
"Hope we can see some more go down"

Discussion

Academics have been alarmed at the rate of fake news and misinformation across social media [28-32]. Initially, social media platforms had been praised for their ability to spread liberal messages during events such as the Arab Spring [22] and during the initial launch of WikiLeaks [33]. False information has been a genuine concern among social media platforms during COVID-19, and Facebook has implemented a new feature that will inform users if they have engaged with false information [34].

One method of counteracting fake news is to be able to detect it rapidly and address it head-on at the time that it occurs. In the specific influencer analysis (in the User Analysis section), there was a lack of an authority figure who was actively combating such misinformation. This study found that a dedicated individual Twitter account set up to spread the conspiracy theory formed a cluster in the network with 408 other Twitter users. This account, at the time of analysis, had managed to send a total of 303 tweets during this specific time period before it was closed down by Twitter. In hindsight, if this account would have been closed down much sooner, this would have halted the spread of this specific conspiracy theory. Moreover, if other users who were sharing humorous content and link baiting the hashtag refrained from tweeting about the topic and instead reported conspiracy-related tweets to Twitter, the hashtag would not have reached trending status on Twitter. As more users began to tweet using the hashtag, the overall visibility increased. Public health authorities may wish to advise citizens against resharing or engaging with misinformation on social media and encourage users to flag them as inappropriate to the social media companies. Many social media platforms provide users with the ability to report inappropriate content.

A further method of counteracting misinformation is to seek the assistance of influential public authorities and bodies such as public figures, government accounts, relevant scientific experts, doctors, or journalists. A further key point to make is that the fight against misinformation should take place on the platform where it arises. This is because people will not go to a website to read the counteracting report, but they will watch a video or a memo voice sent via WhatsApp or posted on a social media platform. Public TV, newspapers, and radio stations could also seek to devote regular programs to counteract fake news by discussing conspiracy theories that were spreading at the time. It could also be argued that it is important to analyze the context of the fake news and why it is spreading. Are people afraid? Does the theory propose a risk? Any content that aims to correct misinformation should aim to dispel people's fears.

This research set out to address four research questions that are now discussed. In regard to identifying how the conspiracy was spreading on Twitter, this article shows that a number of citizens who believed the conspiracy theory were actively tweeting and spreading it (as highlighted in Table 1). A dedicated account that was set up for the sole purpose of spreading the conspiracy theory was identified. We also identified the "humor effect" in the sense that even those users who joined the discussion to mock the conspiracy theory inadvertently drew more attention to it.

In addressing the second research objective, this paper identifies a number of influential online sources that created content aiming to show a link between COVID-19 and 5G (as highlighted in Tables 2 and 3). These consisted of the website *InfoWars*, a commercial organization selling products that protect against electromagnetic fields. A website dedicated to linking 5G to COVID-19 was also identified. Specific YouTube videos and the YouTube domain itself were also found to be influential.

The third research objective was to identify whether people really believed 5G and COVID-19 were linked, as Twitter is known to contain humorous content [16]. It was found that 34.8% (n=81/233) of individual tweets contained views that 5G and COVID-19 were linked. Although it is a low percentage, there are indeed users who genuinely believe COVID-19 and 5G are linked.

In regard to the fourth research objective, this article sought to identify and discuss potential actions public health authorities could take to mitigate the spread of the conspiracy theory. Specifically, this study found that an individual account had been set up to spread the conspiracy theory and was able to attract a following and send out many tweets. Based on our analysis of this conspiracy theory on Twitter, its spread could have been halted if the accounts set up to spread misinformation were taken down faster than they were. Public health authorities should also aim to focus on these types of accounts in combating misinformation during the current COVID-19 pandemic. In addition, an authority figure with a sizeable following could have tweeted messages against the conspiracy theory and urged

other users that the best way to deal with it is to not comment on, retweet, or link bait using the hashtag. This is because when users joined the discussion to dispel, ridicule, or piggyback on the hashtag, the topic was raised to new heights and had increased visibility.

A strength of this study is that it has identified the drivers of the conspiracy theory, the content shared, and the strategies to mitigate the spread of it. Our results are likely to be of international interest during the unfolding COVID-19 pandemic. A further strength of our study is that our methodology can be applied to other conspiracy topics. A limitation of our study is that the Search API can only retrieve data from public facing Twitter accounts. Previous research has noted that certain Twitter topics are likely to contain automated accounts known as "bots" [35]; for instance, in the case of electronic cigarette (e-cigarette) tweets, research has found that social bots could be used to promote new e-cigarette products and spread the idea that they are helpful for smoking cessation [35]. A limitation of our study is that we did not identify social bot accounts; however, influential accounts in our study did not appear to display bot behavior (eg, high number of tweets posted) and appeared to display characteristics of genuine accounts. This could be inferred because certain accounts linked to their profile on other platforms such as YouTube. However, future research could seek to identify the ratio of bots to individual accounts related to conspiracy theories. A further limitation is that our content analysis was conducted on English-language tweets, and further research could seek to examine tweets in other languages. Furthermore, a limitation to our study is that, as we retrieved data using a specific keyword, our data may have excluded tweets from users who tweeted about the conspiracy during this time without using our target keyword or hashtag.

The COVID-19 pandemic has been a serious public health challenge for nations around the world. This study conducted an analysis of a conspiracy theory that threatened to potentially undermine public health efforts. We discussed key users and influential web sources during this time, and discussed potential strategies for combating such dangerous misinformation. The analysis reveals that there was a lack of an authority figure who was actively combating such misinformation, and policymakers should insist in efforts to isolate opinions that are based on fake news if they want to avoid public health damage. Future research could seek to conduct a follow-up analysis of Twitter data as the COVID-19 pandemic evolves.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Top mentioned users.
[DOCX File , 14 KB - jmir_v22i5e19458_app1.docx]

Multimedia Appendix 2
Top replied-to users.
[DOCX File , 14 KB - jmir_v22i5e19458_app2.docx]

Multimedia Appendix 3
Top domains.
[DOCX File , 14 KB - jmir_v22i5e19458_app3.docx]

References

1. Jolly G. Middle East respiratory syndrome coronavirus (MERS-CoV). TIJPH 2016 Dec 31;4(4):351-376. [doi: 10.21522/tijph.2013.04.04.art033]
2. Smith RD. Responding to global infectious disease outbreaks: lessons from SARS on the role of risk perception, communication and management. Soc Sci Med 2006 Dec;63(12):3113-3123 [FREE Full text] [doi: 10.1016/j.socscimed.2006.08.004] [Medline: 16978751]
3. World Health Organization. Middle East respiratory syndrome coronavirus (MERS-CoV) URL: https://www.who.int/emergencies/mers-cov/en/ [accessed 2020-04-15]
4. Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, et al. World Health Organization declares global emergency: a review of the 2019 novel coronavirus (COVID-19). Int J Surg 2020 Apr;76:71-76 [FREE Full text] [doi: 10.1016/j.ijsu.2020.02.034] [Medline: 32112977]
5. Wu F, Zhao S, Yu B, Chen Y, Wang W, Song Z, et al. A new coronavirus associated with human respiratory disease in China. Nature 2020 Mar;579(7798):265-269 [FREE Full text] [doi: 10.1038/s41586-020-2008-3] [Medline: 32015508]
6. Livingston E, Bucher K. Coronavirus disease 2019 (COVID-19) in Italy. JAMA 2020 Mar 17:e. [doi: 10.1001/jama.2020.4344] [Medline: 32181795]
7. Agiwal M, Roy A, Saxena N. Next generation 5G wireless networks: a comprehensive survey. IEEE Commun Surv Tutorials 2016 Feb 19;18(3):1617-1655. [doi: 10.1109/comst.2016.2532458]

8. Quinn B. The Guardian. 2020. Facebook Acts To Halt Far-Right Groups Linking Covid-19 To 5G URL: https://www.theguardian.com/world/2020/apr/12/facebook-acts-to-halt-far-right-groups-linking-covid-19-to-5g [accessed 2020-04-21]
9. Kelion L. BBC News. 2020 Apr 04. Mast fire probe amid 5G coronavirus claims URL: https://www.bbc.com/news/uk-england-52164358# [accessed 2020-04-13]
10. Brewis H. Evening Standard. 2020 Apr 14. Nightingale hospital phone mast attacked as 5G conspiracy theory rages on URL: https://www.standard.co.uk/news/uk/nhs-nightingale-phone-mast-arson-attack-5g-conspiracy-a4414351.html
11. Lewis K. Full Fact. 2020 Mar 31. 5G is not accelerating the spread of the new coronavirus URL: https://fullfact.org/health/5G-not-accelerating-coronavirus/ [accessed 2020-04-14]
12. Shanapinda S. The Conversation. 2020 Apr 07. No, 5G radiation doesn't cause or spread the coronavirus. Saying it does is destructive URL: https://tinyurl.com/uppng7w [accessed 2020-04-14]
13. Waterson J, Hern A. The Guardian. 2020 Apr 06. At least 20 UK phone masts vandalised over false 5G coronavirus claims URL: https://tinyurl.com/v3fv7mh
14. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118. [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
15. Li J, Xu Q, Cuomo R, Purushothaman V, Mackey T. Data mining and content analysis of the Chinese social media platform Weibo during the early COVID-19 outbreak: retrospective observational infoveillance study. JMIR Public Health Surveill 2020 Apr 21;6(2):e18700. [doi: 10.2196/18700] [Medline: 32293582]
16. Ahmed W, Lugovic S. Social media analytics: analysis and visualisation of news diffusion using NodeXL. Online Information Review 2019 Feb 11;43(1):149-160. [doi: 10.1108/oir-03-2018-0093]
17. Smith MA, Rainie L, Himelboim I, Shneiderman B. Pew Research Center. 2014 Feb 20. Mapping Twitter topic networks: from polarized crowds to community clusters URL: https://www.pewresearch.org/internet/2014/02/20/mapping-twitter-topic-networks-from-polarized-crowds-to-community-clusters/
18. Ahmed W, Marin-Gomez X, Vidal-Alaball J. Contextualising the 2019 e-cigarette health scare: insights from Twitter. Int J Environ Res Public Health 2020 Mar 26;17(7):e. [doi: 10.3390/ijerph17072236] [Medline: 32225020]
19. Ahmed W. Public health implications of #ShoutYourAbortion. Public Health 2018 Oct;163:35-41. [doi: 10.1016/j.puhe.2018.06.010] [Medline: 30059806]
20. Shelley M, Krippendorff K. Content analysis: an introduction to its methodology. J Am Stat Assoc 1984 Mar;79(385):240. [doi: 10.2307/2288384]
21. Ahmed W, Bath P, Demartini G. Using Twitter as a data source: an overview of ethical, legal, and methodological challenge. Adv Res Ethics Integrity 2017:79. [doi: 10.1108/s2398-601820180000002004]
22. White DR, Borgatti SP. Betweenness centrality measures for directed graphs. Soc Networks 1994 Oct;16(4):335-346. [doi: 10.1016/0378-8733(94)90015-9]
23. InfoWars. URL: https://www.infowars.com/
24. RayGuardNJ. URL: https://rayguardnj.com/
25. Stillness in the Storm. URL: https://stillnessinthestorm.com/
26. 5G Crisis. URL: https://www.5gcrisis.com/
27. YouTube. Jeff censored! URL: https://www.youtube.com/channel/UCAI2PPsCuPRM6KDbwJvSU9A
28. Downing J, Ahmed W. #MacronLeaks as a “warning shot” for European democracies: challenges to election blackouts presented by social media and election meddling during the 2017 French presidential election. Fr Polit 2019 Jun 13;17(3):257-278. [doi: 10.1057/s41253-019-00090-w]
29. Downing J, Dron R. Tweeting Grenfell: discourse and networks in critical constructions of British Muslim social boundaries on social media. New Media Soc 2019 Jul 26;22(3):449-469. [doi: 10.1177/1461444819864572]
30. Allcott H, Gentzkow M. Social media and fake news in the 2016 election. J Econ Perspect 2017 May;31(2):211-236. [doi: 10.1257/jep.31.2.211]
31. Nussbaum E. The New Yorker. 2017 Jan 23. How jokes won the election URL: https://www.newyorker.com/magazine/2017/01/23/how-jokes-won-the-election
32. Wolfsfeld G, Segev E, Sheafer T. Social Media and the Arab Spring. Int J of Press/Polit 2013 Jan 16;18(2):115-137. [doi: 10.1177/1940161212471716]
33. Sifry ML. WikiLeaks and the Age of Transparency. New York: OR Books; 2011.
34. Tidy J. BBC News. 2020 Apr 16. Coronavirus: Facebook alters virus action after damning misinformation report URL: https://www.bbc.com/news/technology-52309094
35. Allem J, Ferrara E. The importance of debiasing social media data to better understand e-cigarette-related attitudes and behaviors. J Med Internet Res 2016 Aug 09;18(8):e219. [doi: 10.2196/jmir.6185] [Medline: 27507563]

Abbreviations

BCS: betweenness centrality score
COVID-19: coronavirus disease
e-cigarette: electronic cigarette

SARS: severe acute respiratory syndrome
UTC: Coordinated Universal Time

Edited by G Eysenbach; submitted 18.04.20; peer-reviewed by JP Allem, R Zowalla; comments to author 21.04.20; revised version received 22.04.20; accepted 25.04.20; published 06.05.20.

Please cite as:
Ahmed W, Vidal-Alaball J, Downing J, López Seguí F
COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data
J Med Internet Res 2020;22(5):e19458
URL: http://www.jmir.org/2020/5/e19458/
doi:10.2196/19458
PMID:32352383

Viewpoint

Telemedicine During the COVID-19 Pandemic: Experiences From Western China

Zhen Hong[1*], PhD, MD; Nian Li[2*], MD; Dajiang Li[2], MD; Junhua Li[3], BSc; Bing Li[3], BSc; Weixi Xiong[1], MD, PhD; Lu Lu[1], MD; Weimin Li[4], MD; Dong Zhou[1], MD, PhD

[1]Department of Neurology, West China Hospital, Sichuan University, Chengdu, China

[2]Department of Medical Affairs, West China Hospital, Sichuan University, Chengdu, China

[3]Medical Administration, Health Commission of Sichuan Province, Chengdu, China

[4]Department of Pulmonary and Critical Care Medicine, West China Hospital, Sichuan University, Chengdu, China

[*]these authors contributed equally

Corresponding Author:
Dong Zhou, MD, PhD
Department of Neurology
West China Hospital
Sichuan University
No. 37 Guoxue Road
Chengdu, 610041
China
Phone: 86 28 8542 2893
Email: zhoudong66@yahoo.de

Abstract

Disasters and pandemics pose unique challenges to health care delivery. As health care resources continue to be stretched due to the increasing burden of the coronavirus disease (COVID-19) pandemic, telemedicine, including tele-education, may be an effective way to rationally allocate medical resources. During the COVID-19 pandemic, a multimodal telemedicine network in Sichuan Province in Western China was activated immediately after the first outbreak in January 2020. The network synergizes a newly established 5G service, a smartphone app, and an existing telemedicine system. Telemedicine was demonstrated to be feasible, acceptable, and effective in Western China, and allowed for significant improvements in health care outcomes. The success of telemedicine here may be a useful reference for other parts of the world.

(J Med Internet Res 2020;22(5):e19577) doi:10.2196/19577

KEYWORDS

COVID-19; coronavirus disease; medical education; pandemics; teleteaching; tele-education; telemedicine

Introduction

Coronavirus disease (COVID-19) has been declared a pandemic by the World Health Organization (WHO), whose director-general expressed concerns about the “alarming spread and severity” as well as the “alarming levels of inaction” [1]. Although the outbreaks of severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS) were associated with much higher respective case fatality rates (CFRs of 9.6% and 34.4%, respectively), the COVID-19 pandemic has led to more deaths due to the large number of individuals infected. Human-to-human transmission of the SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) virus has become the primary transmission route of the disease [2]. In this context, policymakers need to weigh the risks of inaction and pursue policies and strategies commensurate with the magnitude of the threat.

Pandemics and other public health emergencies typically lead to a surge in demand for medical care, which overwhelms local capabilities. Telemedicine can be broadly defined as the use of telecommunications technologies to provide medical information and services. The benefits of telemedicine in these situations have been well documented [3,4]. Telemedicine can support long-distance clinical care, education, and health administration, and its use has increased dramatically in the past decade.

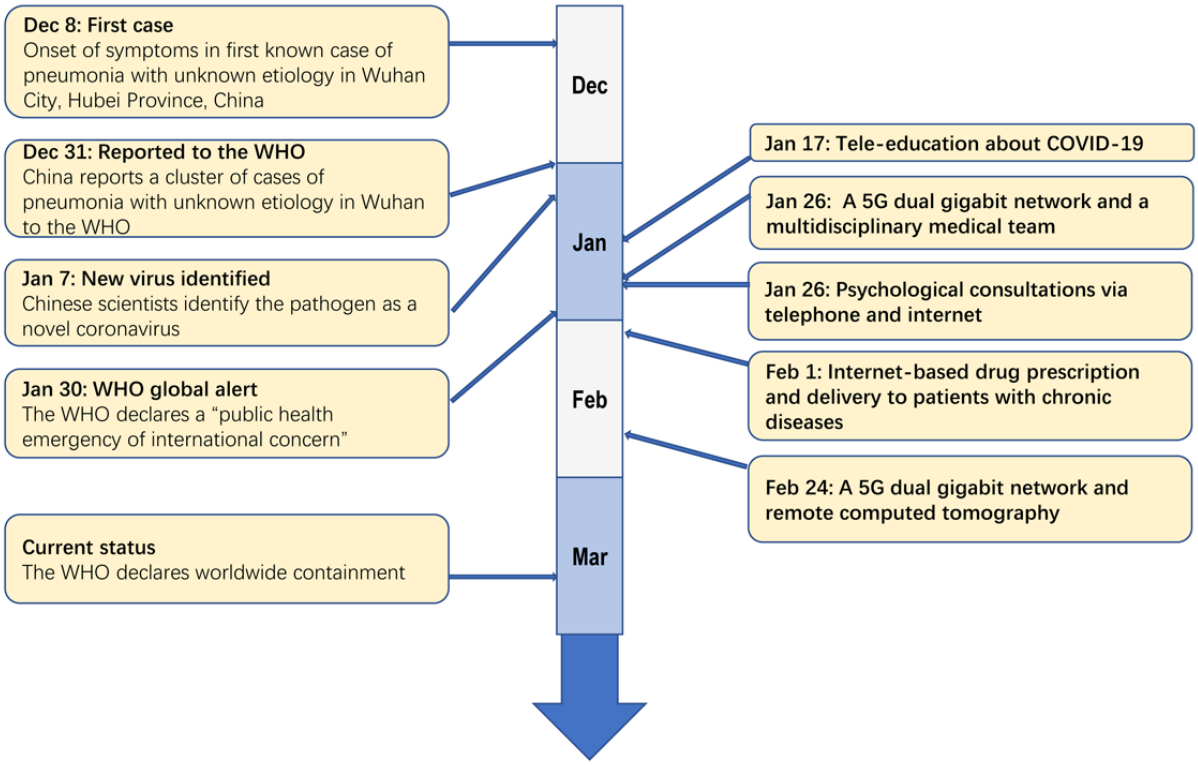
There are a number of potential benefits to implementing telemedicine, including [5]:

- Improved access to information;
- Provision of care not previously deliverable;

- Improved access to services and increasing care delivery;
- Improved professional education;
- Quality control of screening programs;
- Reduced health care costs.

Western China does not have nearly as many economic resources or health care infrastructure as the eastern parts of the country. High-quality medical resources are concentrated in large- and medium-sized cities, and many county- and district-level hospitals face shortages of qualified personnel and inadequate technology for diagnosis and treatment [6]. Here, we share experiences of a multimodal telemedicine network in Sichuan Province in Western China during the COVID-19 pandemic. The network synergizes a newly established 5G service, a smartphone app, and an existing telemedicine system. The Sichuan telemedicine network was activated immediately after the first COVID-19 outbreak in January 2020 (Figure 1). Disaster response funds have been used to pay directly for telehealth care in the short term through a contingency contract with a telehealth network. It capitalized on the remote capabilities of staff and technology and likely played an important role during the pandemic.

Figure 1. Progression of telemedicine in Western China during the coronavirus disease (COVID-19) pandemic. Left: timeline of COVID-19 events since the first case was reported on December 8, 2019. Right: timeline of telemedicine milestones in Sichuan Province of Western China during the COVID-19 pandemic. WHO: World Health Organization.



Tele-Education of Medical Staff About COVID-19

On January 17, 2020, an expert group for rapid response to COVID-19 was established in Sichuan Province; the experts provided tele-education to medical staff at local hospitals via remote consultation networks, portals, and smartphone apps. So far, more than 800,000 person-times have been devoted to training, helping to ensure the implementation of prevention and control measures against epidemics. Training topics include specimen collection methods, laboratory assays of nucleic acids, standardized diagnosis and treatment, prevention and control of hospital infections, personal protection, and medical waste disposal.

A 5G Dual Gigabit Network and a Multidisciplinary Medical Team

On January 26, 2020, the West China Hospital of Sichuan University (WCHSU) launched a new web-based, real-time video telemedicine system for consultations provided by a multidisciplinary team to deal with COVID-19 cases (Figures 2 and 3). The network has focused on groups particularly vulnerable to severe symptoms of COVID-19, including the elderly, pregnant women, children, and patients with chronic health problems. From January 26 to March 12, 2020, 424 remote consultations were conducted for severe and critical COVID-19 patients, whose median age was 64 years (range 35 days to 87 years). Reasons for the consultations included adjustment of a patient's antiviral therapy (75%), management of complications (68%), adjustment of respiratory therapy (55%), or confirmation of COVID-19 diagnosis (15%).

Figure 2. The 5G telemedicine network of Sichuan Province, China, established during the coronavirus disease (COVID-19) pandemic. This system makes use of the newly established China Telecom 5G Dual Gigabit infrastructure and currently covers all designated hospitals for COVID-19 of Sichuan Province (5 provincial-level, 24 municipal-level, and 179 county-level hospitals), with the West China Hospital of Sichuan University (WCHSU) as the central node. The median distance between a spoke hospital and WCHSU is 319 km (range 20 to 1191 km).

Figure 3. Web-based, real-time video telemedicine for consultations provided by a multidisciplinary team to deal with cases of coronavirus disease (COVID-19) in Western China.

This new COVID-19 telemedicine system may increase diagnostic accuracy of difficult cases and improve treatment of severe or critical cases of COVID-19 for the large rural population of Western China at low cost; it has been highly praised by the WHO. This system may help explain why the CFR of COVID-19 is only 0.55% in Sichuan Province, which is much lower than the 4.64% in Hubei Province, China, and the mean rate of the world so far.

A 5G Dual Gigabit Network and Remote Computed Tomography Scanning

On February 24, 2020, radiologists at WCHSU used the 5G Dual Gigabit network to remotely perform computed tomography (CT) on patients with COVID-19 at Ganzi People's Hospital in Ganzi County, Sichuan Province. To our knowledge, this is the first report of remote CT scanning during the COVID-19 pandemic. The experts at WCHSU were able to view the same images and data as the local clinicians at the same time, as well as remotely control the CT equipment at Ganzi People's Hospital. On March 2, 2020, WCHSU clinicians remotely performed CT scanning of a COVID-19 patient at Huangzhou General Hospital in Huanggang City, Hubei Province. So far, 152 patients at hospitals in the telemedicine network have undergone remote CT scanning under the control and guidance of WCHSU physicians.

This marks the transition of telemedicine from the traditional "consultation" mode to a "practical operation" mode, ensuring high-quality CT even in areas with severe shortages of qualified technicians.

COVID-19 Consultations Via Telephone and Internet

WCHSU officially opened a special COVID-19 telephone hotline and smartphone app for online consultations on January 26, 2020, through which medical staff offer free consultations and psychological interventions. By March 23, 2020, 9085 patients had received online consultations or interventions through the app, and 1094 patients had received consultations or interventions from 137 clinicians by telephone. Among these patients, 293 were screened for suspected COVID-19 and followed up.

Internet-Based Drug Prescription and Delivery to Patients With Chronic Diseases

Hospitals are a potential source of COVID-19 cross-infection. To reduce accumulation of people in outpatient clinics during the pandemic, WCHSU and other major Chinese hospitals began to offer online consultations and internet-based drug prescription and delivery service for patients with common and chronic diseases since February 1, 2020. The internet-based services are offered through the hospitals' websites, WeChat accounts, and widely used apps such as Huayitong. By March 23, 2020, 31,905 patients had received prescriptions or medicines through this service. This telemedicine service reduces the number of patient visits, eases overcrowding in outpatient centers, and allays worry among patients with chronic disease.

Strengths, Limitations, and Future Directions

Pandemics pose unique challenges to health care delivery. As our experiences from Western China illustrate, although telemedicine will not solve all challenges, it can provide rapid access to specialists who are unavailable in person. With the rapid development of the internet and smartphone apps, telemedicine has transitioned to a multimodal paradigm, offering greater possibilities and convenience. The successes of telemedicine in Western China may provide a useful reference for other parts of the world. Prerequisites for success include sufficient financial resources, technological infrastructure, and overall arrangement of policymakers.

Telemedicine utilizes information and telecommunications technology to transfer medical information for diagnosis, therapy, and education. Although telemedicine clearly has a wide range of potential benefits, it also has some disadvantages. The main drawbacks of telemedicine are a breakdown in the relationship between health professionals and their patients; a breakdown in the relationship between health professionals; issues concerning the quality of health information; and organizational and bureaucratic difficulties [5]. In the future, with advancements in technology (eg, the application of 5G networks to improve the effect of video transmission) and improvement in the management experience of telemedicine by policymakers, the abovementioned limitations can be minimized, and telemedicine may become a sustainable, mainstream solution for both public health emergencies and routine medicine.

Acknowledgments

This work was supported by the Sichuan Science and Technology Support Program (2020YFS0002).

The data reported in the article are available from the corresponding authors, upon request from qualified investigators.

Authors' Contributions

All authors were involved in drafting the manuscript and agreed to its publication. All authors read and approved their sections of the final manuscript. DZ and WL read and approved all sections of the final manuscript.

Conflicts of Interest

None declared.

References

1. Cucinotta D, Vanelli M. WHO Declares COVID-19 a Pandemic. Acta Biomed 2020 Mar 19;91(1):157-160. [doi: 10.23750/abm.v91i1.9397] [Medline: 32191675]
2. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. JAMA 2020 Feb 24. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
3. Lurie N, Carr BG. The Role of Telehealth in the Medical Response to Disasters. JAMA Intern Med 2018 Jun 01;178(6):745-746. [doi: 10.1001/jamainternmed.2018.1314] [Medline: 29710200]
4. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
5. Hjelm NM. Benefits and drawbacks of telemedicine. J Telemed Telecare 2005 Jan;11(2):60-70. [doi: 10.1258/1357633053499886] [Medline: 15829049]

6. Wang T, Li J, Zhu C, Hong Z, An D, Yang H, et al. Assessment of Utilization and Cost-Effectiveness of Telemedicine Program in Western Regions of China: A 12-Year Study of 249 Hospitals Across 112 Cities. Telemed J E Health 2016 Nov;22(11):909-920. [doi: 10.1089/tmj.2015.0213] [Medline: 27314300]

Abbreviations

CFR: case fatality rate
COVID-19: coronavirus disease
CT: computed tomography
MERS: Middle East respiratory syndrome
SARS: severe acute respiratory syndrome
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
WCHSU: West China Hospital of Sichuan University
WHO: World Health Organization

Edited by G Eysenbach; submitted 23.04.20; peer-reviewed by MA Bahrami, E Da Silva; accepted 29.04.20; published 08.05.20.

Please cite as:
Hong Z, Li N, Li D, Li J, Li B, Xiong W, Lu L, Li W, Zhou D
Telemedicine During the COVID-19 Pandemic: Experiences From Western China
J Med Internet Res 2020;22(5):e19577
URL: http://www.jmir.org/2020/5/e19577/
doi:10.2196/19577
PMID:32349962

Original Paper

Measures Undertaken in China to Avoid COVID-19 Infection: Internet-Based, Cross-Sectional Survey Study

Yu Huang[1*], MSc; Qingqing Wu[1*], MSc; Ping Wang[2], MPH; Yan Xu[1], MSc; Lei Wang[1], MSc; Yusui Zhao[1], MSc; Dingming Yao[1], MSc; Yue Xu[1], MPH; Qiaohong Lv[1], MSc; Shuiyang Xu[1], MPH

[1]Zhejiang Provincial Center for Disease Control and Prevention, Hangzhou, China

[2]Jiashan Prefectural Center for Disease Control and Prevention, Jiashan, China

[*]these authors contributed equally

Corresponding Author:
Shuiyang Xu, MPH
Zhejiang Provincial Center for Disease Control and Prevention
3399 Binsheng Road, Binjiang District
Hangzhou, 310051
China
Phone: 86 87115118
Email: 417286592@qq.com

Abstract

Background: In early 2020, over 80,000 cases of coronavirus disease (COVID-19) were confirmed in China. Public prevention and control measures, along with efforts from all sectors of society, were undertaken to control and eliminate disease transmission.

Objective: This paper describes Chinese citizens' response to the epidemic, the preventive measures they implemented to avoid being infected, and the public strategies that were carried out by the government, health workers, etc. We also discuss the efficacy of these measures in controlling the epidemic in China.

Methods: Information on the responses and behaviors of Chinese citizens were collected through a cross-sectional, internet-based survey using Dingxiang Doctor's public account on WeChat. Information on public strategies implemented by all sectors of society to control the epidemic and data on new COVID-19 cases were collected from the internet, mainly from government websites. Standard descriptive statistics and multivariate logistic regression analyses were conducted to analyze the data.

Results: A total of 10,304 participants responded to the survey, with 10,198 valid responses; 74.1% (n=7557) were female and 25.9% (n=2641) were male. Overall, 98.2% (n=10,013) of participants paid high or very high attention to the epidemic, with WeChat being their main information source (n=9400, 92.2%). Over half the participants (n=5878, 57.7%) were confident that the epidemic could be curbed in China; 92.4% (n=9427) opened windows for ventilation more frequently than usual; 97.9% (n=9986) used masks in public; 95.7% (n=9759) avoided large crowds and stayed at home as much as possible; and 97.9% (n=9988) washed their hands more often than usual. Women were more likely to practice these behaviors than men (P<.001). With a series of strict public control measures, like nationwide health education campaigns, holiday extensions, the Examine and Approve Policy on the resumption of work, close management of working and living quarters, a health QR (Quick Response) code system, community screening, and social distancing policies, the number of new cases have decreased dramatically since February 12, 2020.

Conclusions: The methods employed by Chinese citizens and authorities have effectively curtailed the spread of COVID-19, demonstrating that this pandemic can be brought under control as long as the right measures are taken.

(J Med Internet Res 2020;22(5):e18718) doi:10.2196/18718

KEYWORDS

COVID-19; coronavirus disease; response; strategy; preventive measures; internet-based research; health QR code; outbreak; infectious disease; health education

Introduction

In December 2019, an outbreak of pneumonia associated with the novel coronavirus disease (COVID-19) was reported in Wuhan, Hubei Province, China [1,2]. The timing of the COVID-19 outbreak, prior to the annual Chinese Lunar New Year holiday, coincided with people returning to their family homes, resulting in several billion person-trips made by residents and visitors [3]. In just 30 days, COVID-19 rapidly spread from a single city to the entire country, and from there to other countries around the world [4-7]. On January 30, 2020, the World Health Organization (WHO) declared the outbreak a public health emergency of international concern [8].

COVID-19 is a highly contagious disease. It mainly spreads from person to person through respiratory droplets, similar to the common cold and influenza viruses (ie, through face-to-face contact accompanied by a sneeze or cough). It can also be transmitted through contact with the secretions of infected individuals. The role of fecal–oral transmission is yet to be determined for COVID-19, but it was found to occur during the severe acute respiratory syndrome (SARS) outbreak [9]. The population is generally susceptible to the virus. Based on current epidemiological surveys, the latency period is usually 2-14 days (median 4 days), though longer cases have been noted [10]. COVID-19 is contagious during the latency period. At the time of submission of this paper, targeted antiviral drugs and vaccines were not yet available for COVID-19.

Under such a situation, it is important to provide the public with adequate information on risks and precautions, like the proper use of masks, frequent handwashing, and the avoidance of large gatherings, to control the outbreak. From January 20, 2020, when person-to-person transmission was confirmed and made public in China by Dr Zhong Nanshan [11], a renowned Chinese respiratory scientist, a large-scale, multilevel health education campaign was implemented in China. Information about COVID-19 spread swiftly on the internet, WeChat, microblogs, television, radio, and other media outlets. In rural areas, educational materials such as posters, bulletin boards, banners, and booklets were immediately provided to people with limited access to the internet. Core information included details on the epidemic situation, the government's response to the epidemic, information on COVID-19, and self-protection strategies. Following the recommendations of health authorities, Chinese citizens put in place measures to protect themselves against COVID-19, such as staying at home as much as possible, limiting social contact, and wearing protective masks when in public.

During the early part of the outbreak, genetic analyses conducted in China revealed that the virus was similar to, but distinct from, severe acute respiratory syndrome coronavirus (SARS-CoV). China's emergency management of SARS was heavily criticized in 2003 [12,13], and the Chinese government has improved its epidemic response capacity since then. National and local surveillance systems to prevent and control diseases were established and strengthened; the Field Epidemiology Training Program was initiated to increase the capacity of health workers at different levels; and laboratory capacity, collaboration, and communications with the WHO and the international scientific community were increased and strengthened [14]. The new and potentially serious outbreak of COVID-19 presented an opportunity to evaluate the efficacy of changes in China's emergency management approach.

This paper aimed to describe Chinese citizens' responses to the epidemic, the preventive measures they implemented to avoid being infected, and public strategies carried out by the government, health workers, etc. We also discuss the efficacy of these measures in controlling the epidemic in China.

Methods

Data Collection

Data collection for our study comprised two parts. First, information on personal responses and preventive measures to avoid being infected with COVID-19 were collected through a cross-sectional, internet-based survey. The questionnaire collected sociodemographic data (sex, age, occupation, education level, marriage, etc); sources used to obtain information on COVID-19; the level of attention paid to the epidemic; respondents' confidence in the curbing of the outbreak in China; places that participants had visited during the COVID-19 epidemic period prior to the survey; and preventive measures taken to avoid infection. A 5-point Likert scale (ie, very low=1 to very high=5) was used to evaluate the participants' confidence in curbing the outbreak by China and their level of attention to the COVID-19 epidemic. The questions about preventive measures were as follows:

1. What did you do to protect your families and friends?
2. Did you open windows for ventilation more frequently than usual?
3. Did you wear a mask in public?
4. Did you avoid large crowds and stay at home as much as possible?
5. Did you wash your hands more often than usual?
6. Under the following circumstance, do you wash your hands after touching public goods, toilet use, returning home, coughing/sneezing, or before eating?
7. Do you wash your hands with soap and running water in most cases?
8. Do you cover your coughs/sneezes with tissue or your bent elbow in most cases?

Questions about COVID-19 were based on the latest official report from the WHO, the Chinese Center for Disease Control and Prevention, and scientific literature. The original questionnaire was developed in Chinese by research group members from the Zhejiang Provincial Center for Disease Control and Prevention and refined over two rounds of Delphi method collaboration. A pilot study was conducted to verify its reliability.

Second, information on public strategies carried out by government, health workers, companies, etc, to protect citizens from being infected were sourced from official government websites, national and local health commission websites, national and local centers for disease control and prevention websites, and official papers [15-26]. Data on new COVID-19

cases were collected from the official website of the National Health Commission of the People's Republic of China to describe the epidemic trend in China [15].

Participant Recruitment

Participants were recruited via Dingxiang Doctor, a WeChat public account with 35 million users in China, used to disseminate health knowledge to the general population. A message stating "COVID-19, have you done enough to prevent it?" was created on the site, with a link to the questionnaire. Users who viewed the message could share it through the internet, allowing for news and information to spread quickly to many people. The study period was from January 31 to February 2, 2020.

The study was approved by the Ethics Committee at Zhejiang Provincial Center for Disease Control and Prevention. Informed consent was obtained from all participants before their information was collected.

Statistical Analysis

Data were exported from Dingxiang Doctor to Microsoft Excel and analyzed using SPSS, version 19.0 (IBM Corporation). Standard descriptive statistics were used to summarize the data. Multivariate logistic regression analyses were conducted to explore differences in the practice of preventive behaviors between men and women. *P* values <.05 were considered statistically significant (two-sided). Epidemic trends were depicted visually and main public strategies used in China in response to COVID-19 were chronologically described.

Results

Sociodemographic Characteristics of Respondents

A total of 58,000 Dingxiang Doctor users visited the study page, 17.8% (n=10,304) of whom completed the questionnaire; 106 participants outside of China were excluded and 10,198 responses were included in the study. Table 1 presents participants' sociodemographic characteristics.

Among the participants, 74.1% (n=7557) were female and 25.9% (n=2641) were male; 47.4 (n=4770) were single, divorced, or widowed; 52.6% (n=5364) were married. The proportion of participants in the age groups <30 years, 30-49 years, and ≥50 years were 55.4% (n=5653), 39.8% (n=4059), and 4.8% (n=486), respectively. Most participants had an undergraduate degree (n=7179, 70.4%), while participants with primary or lower, secondary, and postgraduate education comprised 5.1% (n=524), 10.8% (n=1105), and 13.6% (n=1390) of the sample, respectively. A total of 40.1% (n=4089) participants worked in the business and service industry, 17.7% (n=1809) worked in government institutions, 18.1% (n=1844) were students, 6.5% (n=666) were health workers, 7.3% (n=745) were housewives or househusbands, and 10.3% (n=1045) indicated "other" occupational statuses (retired, unemployed, etc).

Table 1. Survey respondents' sociodemographic characteristics (N=10,198).

Characteristic	Respondents, n (%)
Gender	
Male	2641 (25.9)
Female	7557 (74.1)
Age (years)	
<30	5653 (55.4)
30-49	4059 (39.8)
≥50	486 (4.8)
Education	
Primary or less (≤9 years)	524 (5.1)
Secondary (10-12 years)	1105 (10.8)
Undergraduate (13-16 years)	7179 (70.4)
Postgraduate (>16 years)	1390 (13.6)
Occupation	
Staff of government institutions	1809 (17.7)
Staff of business and service industry	4089 (40.1)
Students	1844 (18.1)
Health workers	666 (6.5)
Housewives or househusbands	745 (7.3)
Others (retired, unemployed, etc)	1045 (10.3)
Marital status	
Single/divorced/widowed	4770 (47.4)
Married	5364 (52.6)

Personal Responses to the Epidemic and Preventive Measures Taken to Avoid Infection

Level of Attention to the COVID-19 Epidemic and Information Sources

Using a 5-point Likert scale, most participants indicated that they paid high attention to the epidemic. As seen in Table 2, 78.8% (n=8035) of participants expressed paying very high attention, and 19.4% (n=1978) paid high attention. In terms of information on COVID-19, WeChat was the participants' main source of information (n=9400, 92.2%), followed by news and information apps (n=4529, 44.4%), microblogs (n=4154, 40.7%), television or radio (n=4042, 39.6%), family members, friends, or colleagues (n=2417, 23.7%), websites (n=2167, 21.3%), SMS (n=1110, 10.9%), short video apps (n=1606, 15.8%), community advocacy (n=1069, 10.5%), and paper-based media (n=671, 6.6%).

Table 2. Chinese citizens' level of attention to the coronavirus disease (COVID-19) epidemic and their main information sources.

Variables	Total responses (N=10,198), n (%)
How do you rate your attention to the COVID-19 epidemic?	
Very high	8035 (78.8)
High	1978 (19.4)
Neutral	170 (1.7)
Low	11 (0.1)
Very low	4 (0.04)
Where did you get information on the epidemic?	
WeChat	9400 (92.2)
News and information applications	4529 (44.4)
Microblogs	4154 (40.7)
Television or radio	4042 (39.6)
Family members/friends/colleagues	2417 (23.7)
Websites	2167 (21.3)
Short video applications	1606 (15.8)
SMS	1110 (10.9)
Community advocacy	1069 (10.5)
Paper media (newspaper, magazine, etc)	671 (6.6)

Preventive Measures to Avoid COVID-19 Infection

Most participants followed the recommendations of the public health authority. Table 3 shows that 92.4% (n=9427) of participants opened windows for ventilation more frequently than usual, 97.9% (n=9986) used masks in public, 95.7% (n=9759) avoided large crowds and stayed at home as much as possible, and 97.9% (n=9988) washed hands more often than usual. As for circumstances under which respondents felt inclined to wash their hands, 88.1% (n=8980) of participants washed their hands after touching public goods, 94.8% (n=9665) after toilet use, 91.5% (n=9331) after returning home, 66.5% (n=6782) after coughing or sneezing, and 87.5% (n=8923) before eating. In terms of how participants washed their hands, 86.7% used soap and running water in most cases. However, only 57.8% (n=5889) of participants covered their coughs and sneezes with tissue or a bent elbow. Furthermore, we found that most participants had tried to influence their families and friends: 84.8% (n=8652) told families and friends to avoid large gatherings, 88.6% (n=9037) asked them to stay at home as much as possible, 82.0% (n=8269) shared information on the epidemic with them, and 85.5% (n=8722) told them to wear a mask in public.

Gender differences existed in the practice of all preventive behaviors, with women being more compliant with all hygiene measures than men. Except for differences in "washing hands after coughing/sneezing or before eating" and "told friends and family to avoid large gatherings," all differences were statistically significant ($P<.05$).

Table 3. Preventive measures taken by participants to avoid being infected with coronavirus disease (COVID-19), grouped by gender.

Variable	Gender		Total, n (%)	*P* value[a]
	Male, n (%)	Female, n (%)		
What did you do to protect your family or friends?				
Told them to avoid large gatherings	2194 (83.1)	6458 (85.5)	8652 (84.8)	.07
Persuaded them to stay at home as much as possible	2276 (86.2)	6761 (89.5)	9037 (88.6)	<.001
Shared epidemic information with them	2062 (78.1)	6207 (82.1)	8269 (82.0)	<.001
Told them to use a mask in public	2205 (83.5)	6517 (86.2)	8722 (85.5)	<.001
Did you open windows for ventilation more frequently than usual?				.003
Yes	2394 (90.6)	7033 (93.1)	9427 (92.4)	
No	247 (9.4)	524 (6.9)	771 (7.6)	
Did you wear a mask in public?				.001
Yes	2563 (97.0)	7423 (98.2)	9986 (97.9)	
No	78 (3.0)	134 (1.8)	212 (2.1)	
Did you avoid large crowds and stayed at home as much as possible?				<.001
Yes	2488 (94.2)	7271 (96.2)	9759 (95.7)	
No	153 (5.8)	286 (3.8)	439 (4.3)	
Did you wash your hands more often than usual?				<.001
Yes	2557 (96.8)	7431 (98.3)	9988 (97.9)	
No	84 (3.2)	126 (1.7)	210 (2.1)	
Under the following circumstance, do you wash your hands?				
After touching public goods	2262 (85.7)	6718 (88.9)	8980 (88.1)	.001
After toilet use	2424 (91.8)	7241 (95.8)	9665 (94.8)	<.001
After returning home	2330 (88.2)	7001 (92.6)	9331 (91.5)	<.001
After coughing/sneezing	1706 (64.6)	5076 (67.2)	6782 (66.5)	.06
Before eating	2279 (86.3)	6644 (87.9)	8923 (87.5)	.11
Do you wash your hands with soap and running water in most cases?				<.001
Yes	2196 (83.2)	6646 (87.9)	8842 (86.7)	
No	445 (16.8)	911 (12.1)	1356 (13.3)	
Do you cover your coughs/sneezes with tissue or a bent elbow in most cases?				.01
Yes	1451 (54.9)	4438 (58.7)	5889 (57.8)	
No	1190 (45.1)	3119 (41.3)	4309 (42.3)	

[a]Multivariate logistic regression, adjusted for age, education level, occupation, marital status, and region.

Places Participants Had Visited During the COVID-19 Outbreak (Before the Survey)

From January 20, 2020, when person-to-person transmission was confirmed and made public in China, social distancing policies were advocated for among the public. As shown in Table 4, the most common place visited by participants before the survey were supermarkets or malls (n=5109, 50.1%); 26.9% (n=2648) of participants always stayed at home, 17.2% (n=1755) of participants went out to gather with friends or family members, 14.0% (n=1427) visited farmers' markets, 11.0% (n=1120) used public transport, and 8.8% (n=898) went to their workplaces.

Table 4. Places that participants had been to during the coronavirus disease (COVID-19) epidemic period prior to taking the survey.

Place	Total responses (N=10,198), n (%)
Supermarket or shopping mall	5109 (50.1)
Always at home	2648 (26.0)
Gathering with friends or family members	1755 (17.2)
Farmer's market	1427 (14.0)
Public transport areas	1120 (11.0)
Workplace	898 (8.8)

Participants' Confidence in Curbing the COVID-19 Epidemic

Over half the participants were confident that the COVID-19 epidemic would be curbed in China, among whom 29.6% (n=3014) participants were strongly confident, and 28.1% (n=2864) were confident. A total of 30.1% (n=3073) participants were neutral, 10.0% (n=1022) lacked confidence, and 2.2% (n=225) strongly lacked confidence (Table 5).

Table 5. Participants' confidence in curbing the COVID-19 epidemic in China.

Confidence level	Total responses (N=10,198), n (%)
Strongly confident	3014 (29.6)
Confident	2864 (28.1)
Neutral	3073 (30.1)
Lack of confidence	1022 (10.0)
Strong lack of confidence	225 (2.2)

Main Public Strategies and Epidemic Trend of COVID-19 in China

Since January 20, 2020, a series of strategies were implemented to curb the spread of the virus to the wider community. Table 6 describes the main public strategies implemented by the government, health workers, factories, companies, etc, in response to the epidemic. In addition to active contact tracing, isolation and quarantine policies, a health education campaign, extension of the Chinese Lunar New Year holiday, social distancing policies, an Examine and Approve Policy on the resumption of work, close management of working and living spaces, and a health QR (Quick Response) code system, among others, were implemented.

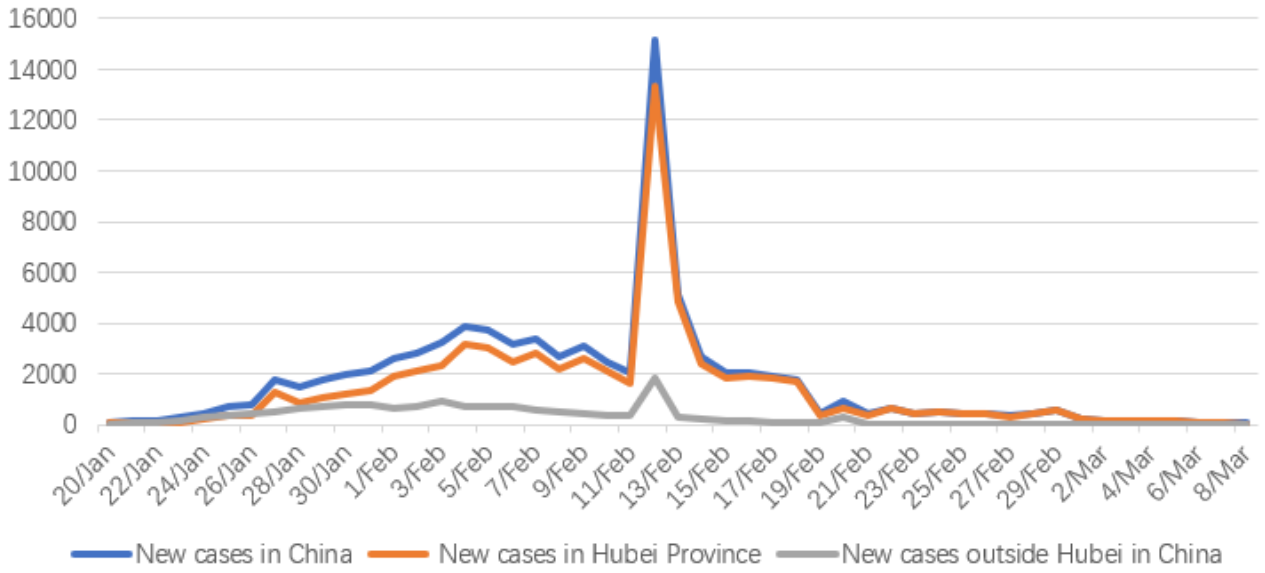
Figure 1 shows the epidemic trend of COVID-19 in China. With the approach of the Lunar New Year holiday, mass population movement provided opportunity for the spread of COVID-19, and new cases emerged and increased in cities outside of Wuhan. With social distancing and a range of accompanying epidemic control measures, the number of daily new cases decreased continually after February 12.

Table 6. Main public strategies that responded to coronavirus disease 2019 (COVID-19) epidemic in China.

Date	Main strategies implemented by the government, health workers, factories, companies, and media
January 20	Person-to-person transmission is officially announced to the public. Nationwide health education campaign is initiated. People are encouraged to stay at home, avoid gatherings, wear protective masks when they need to move in public, etc.
January 23	The government extends Chinese Lunar New Year holiday. Museums, libraries, shopping malls, etc, are closed. Large public events are canceled or postponed. Chinese authorities place a lockdown on Wuhan, the epicenter of COVID-19, and traffic in Wuhan and cities across Hubei Province is restricted and monitored. Transportation is subsequently restricted at a national level.
January 24	The first batch of medical teams from outside Hubei arrive in Wuhan. More medical staff continue to arrive, totaling over 40,000 individuals. Chinese companies are ordered to build a 1000-bed hospital within 10 days. Work on a second facility with 1300 beds follows 2 days later. To plug the shortage of protective suits, masks, and other medical supplies, Chinese manufacturers from various industries are mobilized, including those that normally manufacture cars and cellphones.
January 27	The Examine and Approve Policy on resumption of work is initiated. Factories, companies, etc, begin to collect the travel history and health status of staff members.
February 3	Psychological service is provided nationally. A training program is set up on the National Health Commission website for continuing medical education [20] to share the latest knowledge on prevention, diagnosis, and treatment of COVID-19, to enhance the capacity of health workers.
February 4	Close management of communities, villages, and workplaces to curb COVID-19 begins in Hangzhou and is later implemented nationally. Community screening is initiated.
February 10	China urges efforts to ensure orderly resumption of work to provide sufficient material support for epidemic control. The health status of workers is examined and recorded before they resume work and monitored twice a day during work.
February 11	A health QR[a] code system is developed by Alibaba Company, and implemented to control the spread of COVID-19 in Hangzhou. People traveling to Hangzhou must report their travel history and health conditions in advance online and are issued with green, yellow, or red QR codes, based on the information they had provided. A green code holder, rated as having little chance of being infected, can visit public areas and take public transport normally after taking their temperatures. Those with yellow and red codes, however, must be quarantined for 14 days and report their health information every day, or be sent to a hospital if necessary, before they may travel. Subsequently, the system is implemented nationally.
February 24	With the decrease in new cases, shopping malls, libraries, museums, etc, are reopened in succession.
February 28	The government publicly announces that the COVID-19 epidemic in Wuhan is controllable and is heading in a good direction.

[a]QR: Quick Response.

Figure 1. Daily new cases of coronavirus disease (COVID-19) in China since January 20, 2020.



Discussion

Principal Findings

This is the first large-scale, nationwide study to report on the comprehensive preventive measures taken by Chinese citizens and strategies implemented by the government, health workers, factories, etc, to combat COVID-19 at a national level. We found that Chinese citizens responded quite well to the COVID-19 epidemic, by strictly following the recommendations of health authorities. The frequency of taking preventive

measures by participants was high, with 92.4% of participants opening windows for ventilation more often than usual, 99.4% of participants using masks in public, 95.7% of participants avoiding large crowds and staying at home as much as possible, and 97.9% of participants washing hands more often than usual. However, responses varied among participants, and women showed higher compliance with all hygiene measures than men—a phenomenon that was also found in other studies [27-30]. Furthermore, fewer males than females participated in our survey (25% vs 75%); another internet-based study on Zika also had a similar gender distribution in terms of participants (30% male vs 70% female) [30]. It may be that women are generally more concerned about their health than men, so they are more willing to accept official recommendations and participate in health-related surveys.

Compared with measures taken to prevent the transmission of SARS in 2003 [30], we found a much higher frequency of masks usage (99.4% vs 68.7%), although we also found a lower frequency of covering one's mouth when sneezing or coughing (57.8% vs 70.6%), and a lower frequency of hand washing after sneezing or coughing (66.5% vs 75.9%). The frequency of washing hands with soap was similar between participants of the two studies. Our findings suggest that although the health education campaign was quite effective in China, further education on practicing respiratory hygiene is needed, as sneezing and coughing are an important transmission route for COVID-19. Moreover, gender should be taken into consideration when planning and implementing health promotion and education programs.

WeChat, the largest standalone social media mobile app in China, was found to be one of the main sources of information on the COVID-19 epidemic for over 90% participants. In China, over 95% of adults own a mobile phone and over 1 billion access WeChat at least once a day. The app has become an integral part of Chinese daily life, with citizens using it to send messages, share updates, and access the latest news from the government, celebrities, enterprises, and so on [31]. Our study further confirmed that WeChat has become an ideal platform for delivering health-related information. Other studies from China also found WeChat to be an effective health behavior intervention platform [31-33].

Our findings showed that over 80% participants had tried to influence their families and friends by telling them to avoid large gatherings and stay at home as much as possible, sharing epidemic information with them, and telling them to use masks when necessary. This finding supports previous findings that participants who lived with their families were more likely to use a higher number of precautionary measures than those who did not live with their families during the SARS outbreak in 2003 [30].

Innovative measures combined with traditional strategies have facilitated China's response to COVID-19. Traditional strategies, including active contact tracing, isolation and quarantine, a series of social distancing policies, and community containment were reinforced at a national level. China's emergency management of SARS was heavily criticized in 2003 [12,13]; however, rapid and decisive strategies and the continuing decline of daily new cases of COVID-19 have earned favorable comment [14]. In addition to traditional measures, big data and mobile internet technologies were used to prevent the spread. On February 11, 2020, a unified health QR code system was promoted, first in Hangzhou, and then nationwide. The QR codes, generated with the mobile app Alipay, were based on users' movements over a 14-day period, including whether users had been to virus-hit areas and had contact with confirmed or suspected cases. Instead of completing a health report form, Chinese citizens can now show the QR code when required; this is a much more time efficient approach to reducing the risk of virus transmission. It has also helped to facilitate traffic and speed up the resumption of work and production, ultimately promoting regional economic recovery [18]. Further, a training program was set up on the National Health Commission website for continuing medical education to share the latest knowledge on prevention, diagnosis, and treatment of COVID-19 to enhance the capacity of health workers around China [20]. Due to restrictions on gatherings during the epidemic and continuous deepening of knowledge on COVID-19, online education is an appropriate way to empower health workers.

Limitations

This study has some limitations. First, selection bias might exist. Intrinsic limits, such as partial coverage of the population and missing responses from interviewees, are also present in our study, like other internet-based surveys [34,35]. Participants in our study were younger and more educated compared to the whole population. However, this allowed for the participation of a large group of geographically disparate participants at one discrete time point in a cost-effective and time-efficient manner. For population studies on infectious diseases, like SARS and COVID-19, internet-based surveys are the most appropriate method for data collection, since it avoids transmission of the virus during face-to-face investigation. Second, the number of daily new cases used to describe the trend of the COVID-19 epidemic in China is likely higher due to insufficient testing capacity and difficulties in identifying and counting mild and asymptomatic cases. However, the decline in new cases creates capacity for daily testing. Third, public strategies described in our study represented the main measures taken by the government, health workers, companies, and factories, and we may have missed some details due to restrictions on paper length.

Conclusions

Our study provided a broad description of preventive measures taken by Chinese citizens to avoid being infected with the novel coronavirus and the main strategies implemented by the government, health workers, factories, companies, and media to protect citizens. Chinese citizens followed the hygiene recommendations of health authorities very well; however, further education on practicing respiratory hygiene is still needed in China. In addition to traditional response strategies, an innovative health QR code system based on big data and mobile internet technologies has helped to prevent the spread of COVID-19 and facilitate the resumption of work and production. Online education has enhanced the capacity of health workers on the prevention, diagnosis, and treatment of COVID-19 in a

safe and efficient way. With daily decreases in the number of new cases, we conclude that Chinese measures to curb the spread of COVID-19 has been effective. Our findings suggest that an epidemic of COVID-19 can be brought under control as long as the right measures are taken.

Acknowledgments

We are grateful to the staff of Ding Xiang Doctor WeChat platform for their help in conducting the study and for the support by Zhejiang Medicine and Health Technology project (2018KY335).

Conflicts of Interest

None declared.

References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med 2020 Feb 20;382(8):727-733. [doi: 10.1056/nejmoa2001017]
2. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. N Engl J Med 2020 Mar 26;382(13):1199-1207 [FREE Full text] [doi: 10.1056/NEJMoa2001316] [Medline: 31995857]
3. Ministry of Transport of the People's Republic of China. 2020 Jan 09. Big data! The travel volume predictions during Lunar New Year holiday in 2020 URL: http://www.mot.gov.cn/fenxigongbao/yunlifenxi/202001/t20200109_3322161.html [accessed 2020-02-15]
4. Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. [The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China]. Chinese Journal of Epidemiology 2020 Feb 17;41(2):145-151 [FREE Full text] [doi: 10.3760/cma.j.issn.0254-6450.2020.02.003] [Medline: 32064853]
5. Lim J, Jeon S, Shin HY, Kim MJ, Seong YM, Lee WJ, et al. Case of the Index Patient Who Caused Tertiary Transmission of COVID-19 Infection in Korea: the Application of Lopinavir/Ritonavir for the Treatment of COVID-19 Infected Pneumonia Monitored by Quantitative RT-PCR. J Korean Med Sci 2020 Feb 17;35(6):e79 [FREE Full text] [doi: 10.3346/jkms.2020.35.e79] [Medline: 32056407]
6. Spina S, Marrazzo F, Migliari M, Stucchi R, Sforza A, Fumagalli R. The response of Milan's Emergency Medical System to the COVID-19 outbreak in Italy. The Lancet 2020 Mar;395(10227):e49-e50. [doi: 10.1016/s0140-6736(20)30493-1]
7. Jernigan DB, CDC COVID-19 Response Team. Update: Public Health Response to the Coronavirus Disease 2019 Outbreak - United States, February 24, 2020. MMWR Morb Mortal Wkly Rep 2020 Feb 28;69(8):216-219 [FREE Full text] [doi: 10.15585/mmwr.mm6908e1] [Medline: 32106216]
8. World Health Organization. Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). 2020 Presented at: Emergency Committee / WHO; Jan 30, 2020; Geneva URL: https://tinyurl.com/rjdtx2k [doi: 10.2807/1560-7917.es.2020.25.5.200131e]
9. Lee SH. The SARS epidemic in Hong Kong. J Epidemiol Community Health 2003 Sep 01;57(9):652-654 [FREE Full text] [doi: 10.1136/jech.57.9.652] [Medline: 12933765]
10. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. 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] [Medline: 32109013]
11. XinhuaNet. 2020 Jan 20. China confirms human-to-human transmission of 2019-nCoV, infection of medical staff URL: http://www.xinhuanet.com/english/2020-01/20/c_138721785.htm [accessed 2020-03-13]
12. Heymann DL. The international response to the outbreak of SARS in 2003. Philos Trans R Soc Lond B Biol Sci 2004 Jul 29;359(1447):1127-1129 [FREE Full text] [doi: 10.1098/rstb.2004.1484] [Medline: 15306399]
13. Heymann D. SARS and emerging infectious diseases: a challenge to place global solidarity above national sovereignty. Ann Acad Med Singapore 2006 May;35(5):350-353 [FREE Full text] [Medline: 16830003]
14. Nkengasong J. Author Correction: China's response to a novel coronavirus stands in stark contrast to the 2002 SARS outbreak response. Nat Med 2020 Mar 5;26(3):441-441 [FREE Full text] [doi: 10.1038/s41591-020-0816-5] [Medline: 32139890]
15. National Health Commission of the People's Republic of China. Updates on the epidemic. URL: http://www.nhc.gov.cn/xcs/yqtb/list_gzbd.shtml [accessed 2020-03-13]
16. The State Council of the People's Republic of China. 2020 Jan 27. The State Council's announcement on extending the Lunar New Year Holiday in 2020 URL: http://www.gov.cn/zhengce/content/2020-01/27/content_5472352.htm [accessed 2020-03-13]
17. National Health Commission of the People's Republic of China. 2020 Mar 13. Updates on prevention and control of COVID-19 URL: http://www.nhc.gov.cn/xcs/fkdt/list_gzbd.shtml [accessed 2020-03-13]

18. XinhuaNet. 2020 Mar 05. Health QR code services expand to Yangtze River Delta URL: http://www.xinhuanet.com/english/2020-03/05/c_138846936.htm [accessed 2020-03-13]
19. Ezhejiang.gov. 2020 Feb 13. Zhejiang to promote health QR code system URL: http://www.ezhejiang.gov.cn/2020-02/13/c_452063.htm [accessed 2020-03-13]
20. Chinese Continuing Medical Education. 2020 Feb 03. Training program on COVID-19 URL: http://www.ncme.org.cn/ncmeportal/#/coronavirus/courses [accessed 2020-03-13]
21. CCTV. 2020 Feb 28. Wuhan's war on COVID-19: How China mobilizes the whole country to contain the virus URL: http://www.newsgd.com/news/2020-02/28/content_190450320.htm [accessed 2020-03-13]
22. China.org.cn. 2020 Jan 27. Shanghai postpones resumption of work, school openings amid coronavirus outbreak URL: http://www.china.org.cn/china/Off_the_Wire/2020-01/27/content_75652766.htm [accessed 2020-03-13]
23. The People's Government of Beijing Municipality. 2020 Feb 01. Companies can flexibly arrange jobs during the novel coronavirus epidemic except for those absolutely necessary for the normal operation of cities URL: http://www.gov.cn/xinwen/2020-02/01/content_5473522.htm [accessed 2020-03-13]
24. The People's Government of Hangzhou Municipality. 2020 Jan 27. The announcement on postponing the reoperation date of companies and the reopening date of schools URL: http://www.hangzhou.gov.cn/art/2020/2/9/art_1256295_41893739.html [accessed 2020-03-13]
25. XinhuaNet. 2020 Feb 10. China urges efforts to ensure orderly resumption of work URL: http://www.xinhuanet.com/english/2020-02/10/c_138768851.htm [accessed 2020-03-13]
26. The People's Government of Hangzhou Municipality. 2020 Feb 10. The announcement of 10 preventive measures to fight COVID-19 in Hangzhou URL: http://baijiahao.baidu.com/s?id=1658369539585023074&wfr=spider&for=pc [accessed 2020-03-13]
27. Hernandez EM, Margolis R, Hummer RA. Educational and Gender Differences in Health Behavior Changes After a Gateway Diagnosis. J Aging Health 2018 Mar 08;30(3):342-364 [FREE Full text] [doi: 10.1177/0898264316678756] [Medline: 27940641]
28. Gu H, Jiang Z, Chen B, Zhang JM, Wang Z, Wang X, et al. Knowledge, Attitudes, and Practices Regarding Avian Influenza A (H7N9) Among Mobile Phone Users: A Survey in Zhejiang Province, China. JMIR Mhealth Uhealth 2015 Feb 04;3(1):e15 [FREE Full text] [doi: 10.2196/mhealth.3394] [Medline: 25653213]
29. Huang Y, Xu S, Wang L, Zhao Y, Liu H, Yao D, et al. Knowledge, Attitudes, and Practices Regarding Zika: Paper- and Internet-Based Survey in Zhejiang, China. JMIR Public Health Surveill 2017 Oct 30;3(4):e81 [FREE Full text] [doi: 10.2196/publichealth.7663] [Medline: 29084711]
30. So WK, Chan SS, Lee AC, Tiwari AF. The knowledge level and precautionary measures taken by older adults during the SARS outbreak in Hong Kong. Int J Nurs Stud 2004 Nov;41(8):901-909 [FREE Full text] [doi: 10.1016/j.ijnurstu.2004.04.004] [Medline: 15476763]
31. Shen L, Wang S, Chen W, Fu Q, Evans R, Lan F, et al. Understanding the Function Constitution and Influence Factors on Communication for the WeChat Official Account of Top Tertiary Hospitals in China: Cross-Sectional Study. J Med Internet Res 2019 Dec 09;21(12):e13025 [FREE Full text] [doi: 10.2196/13025] [Medline: 31815674]
32. Zhang W, Li Z, Li Z. WeChat as a Platform for Problem-Based Learning in a Dental Practical Clerkship: Feasibility Study. J Med Internet Res 2019 Mar 19;21(3):e12127 [FREE Full text] [doi: 10.2196/12127] [Medline: 30888328]
33. He C, Wu S, Zhao Y, Li Z, Zhang Y, Le J, et al. Social Media-Promoted Weight Loss Among an Occupational Population: Cohort Study Using a WeChat Mobile Phone App-Based Campaign. J Med Internet Res 2017 Oct 23;19(10):e357 [FREE Full text] [doi: 10.2196/jmir.7861] [Medline: 29061555]
34. Hodder RK, Wolfenden L. Comparison of online and paper survey participation rates in a child health survey by parents of secondary school students. Aust N Z J Public Health 2017 Oct 29;41(5):547-548. [doi: 10.1111/1753-6405.12682] [Medline: 28664641]
35. Assi S, Thomas J, Haffar M, Osselton D. Exploring Consumer and Patient Knowledge, Behavior, and Attitude Toward Medicinal and Lifestyle Products Purchased From the Internet: A Web-Based Survey. JMIR Public Health Surveill 2016 Jul 18;2(2):e34 [FREE Full text] [doi: 10.2196/publichealth.5390] [Medline: 27430264]

Abbreviations

COVID-19: coronavirus disease 2019
QR: Quick Response
SARS: severe acute respiratory syndrome
SARS-CoV: severe acute respiratory syndrome coronavirus
WHO: World Health Organization

Edited by G Eysenbach; submitted 16.03.20; peer-reviewed by L Li, A Tsanas; comments to author 30.03.20; revised version received 09.04.20; accepted 29.04.20; published 12.05.20.

Please cite as:
Huang Y, Wu Q, Wang P, Xu Y, Wang L, Zhao Y, Yao D, Xu Y, Lv Q, Xu S
Measures Undertaken in China to Avoid COVID-19 Infection: Internet-Based, Cross-Sectional Survey Study
J Med Internet Res 2020;22(5):e18718
URL: http://www.jmir.org/2020/5/e18718/
doi:10.2196/18718
PMID:32396516

Original Paper

Analysis of Public Perception of the Israeli Government's Early Emergency Instructions Regarding COVID-19: Online Survey Study

Anat Gesser-Edelsburg[1,2], PhD; Ricky Cohen[1,2], BSN, MHA; Rana Hijazi[1,2], BSc, MPH; Nour Abed Elhadi Shahbari[1,2], BSc, MHA

[1]School of Public Health, University of Haifa, Haifa, Israel

[2]The Health and Risk Communication Research Center, University of Haifa, Haifa, Israel

Corresponding Author:
Anat Gesser-Edelsburg, PhD
School of Public Health
University of Haifa
199 Aba Khoushy Ave Mount Carmel
Haifa, 3498838
Israel
Phone: 972 544243530
Email: ageser@univ.haifa.ac.il

Abstract

Background: On March 11, 2020, the World Health Organization (WHO) officially declared coronavirus disease (COVID-19) to be a pandemic. This posed challenges to many countries, prominent among which is communication with the public to gain their cooperation. Israel faces different challenges from other countries in its management of the COVID-19 crisis because it is in the midst of a deep constitutional crisis.

Objective: The objective of this paper was to examine the response of the Israeli public to the government's emergency instructions regarding the pandemic in terms of correlations between overall risk perception and crisis management; overall risk perception and economic threat perception; crisis management and compliance with behavioral guidelines; and crisis management and economic threat perception. We also made comparisons between crisis management and spokesperson credibility and between crisis management and the credibility of information sources.

Methods: The sample was established using an online survey that enabled rapid and effective distribution of an online questionnaire during the COVID-19 crisis. The self-selection online survey method of nonprobability sampling was used to recruit participants (N=1056) through social network posts asking the general public (aged ≥18 years) to answer the survey.

Results: Participants aged ≥65 years perceived higher personal risk compared to those aged 18-30 years (mean difference 0.33, 95% CI 0.04-0.61) and those aged 46-64 years (mean difference 0.38, 95% CI 0.12-0.64). Significant correlations were found between overall risk perception and attitudes toward crisis management (r=0.19, P<.001), overall risk perception and economic threat perception (r=0.22, P<.001), attitudes toward crisis management and compliance with behavioral guidelines (r=0.15, P<.001), and attitudes toward crisis management and economic threat perception (r=–0.15, P<.001). Participants who perceived that the prime minister was the most credible spokesperson evaluated the crisis management significantly higher than all other groups. The crisis management was evaluated significantly lower by participants who stated that infectious disease specialists were the most credible spokespersons. Participants for whom the Ministry of Health website was the most credible source of information evaluated the crisis management higher than all other groups. Participants for whom scientific articles were the most credible source of information evaluated the crisis management lower than those who perceived that the WHO/Centers for Disease Control and Prevention websites or Ministry of Health/hospital websites and health care workers were the most credible.

Conclusions: The higher the public trust and evaluation of crisis management, the greater the compliance of the public with guidelines. It was also found that crisis management and information cannot be approached in the same way for the overall public. Furthermore, unlike other epidemics, the COVID-19 crisis has widespread economic and social consequences; therefore, it is impossible to focus only on health risks without communicating economic and social risks as well.

(J Med Internet Res 2020;22(5):e19370) doi:10.2196/19370

KEYWORDS

covid-19; risk; perception; crisis management; economical threat; compliance to guidelines; spokesperson credibility; source of information credibility; online survey; public health; pandemic

Introduction

On January 30, 2020, the World Health Organization (WHO) declared the coronavirus disease (COVID-19) outbreak to be a Public Health Emergency of International Concern [1]. The virus continued to spread and cross international borders, and on March 11, 2020, the WHO officially declared COVID-19 a pandemic [2,3]. This pandemic has posed challenges to many countries; prominent among these is communicating with the public to gain their cooperation [4,5].

Despite awareness of the centrality and importance of emerging infectious disease communication, many communication failures have occurred globally and locally surrounding successive epidemic outbreaks, such as bovine spongiform encephalopathy ("mad cow disease"), severe acute respiratory syndrome (SARS), H1N1, Zika virus, and Ebola virus [6,7]. Studies have indicated that mistrust in authorities, lack of information transparency, and failure to customize information to different subpopulations are the main reasons for the failures of emerging infectious disease communication over the years [8,9]. Public trust in government institutions and leaders is essential in any country that seeks to impose authority and maintain public order. If citizens of a country do not place trust in authority, the political, economic and social stability of that country are liable to be harmed [10]. Research shows that the degree of trust in the health system has a major impact on public willingness to receive health instructions and to seek out offered services [11,12]. Lack of cooperation and low levels of trust can cause the public to distance itself from the health system, thus exposing individuals and society to health complications [13]. Trusting an institution implies that individuals believe the entity is generally competent, is able to fulfill its obligations toward its constituents, and acts in responsible ways [11]. In decision-making during health crises, individuals must trust the information they receive, and they must trust the organizations and their spokespersons who communicate the information [14,15]. Conveying information and communicating risk to the public during the COVID-19 crisis are becoming complicated issues because of the ongoing uncertainty surrounding the source and spread of the virus and the absence of a vaccine [16-20].

During a health crisis, policymakers must state uncertainty and share all existing information with the public while addressing and customizing the information to different populations to earn public trust [9] and not lose cooperation, such as in the Ebola case in the US [21] and the polio crisis in Israel [22].

On the individual level, the WHO and health authorities have issued instructions to the public on how to avoid contracting COVID-19 [3]. On the state level, COVID-19 has created national emergencies. Each country has established its own policy to manage the epidemic [23,24]. Measures range from strident to lax: China enacted extreme measures, including a general curfew, shutting down air and land travel, prohibiting public gatherings, building hospitals for patients with COVID-19, and hiring health care personnel; South Korea enacted diverse measures, including a combination of monitoring and careful screening based on increasing the number of tests and using electronic surveillance systems to monitor patients [25-27]. Some countries, such as the United States and the United Kingdom, changed their approaches from lenient to stringent in the middle of the global crisis [28].

Israel is one of the countries that responded to the crisis early. On February 27, the first COVID-19 case was confirmed in Israel. After that, schools were shut down, gatherings were prohibited, electronic surveillance measures were introduced by the government to monitor citizens, and emergency regulations were enacted, including imposing a curfew and allowing people to exit their homes only for critical reasons [29].

As of April 12, 2020, the number of confirmed COVID-19 cases in Israel was 10,878, which is an infection rate of 1257 cases per million citizens. There were 103 deaths, which is 12 per million citizens. Relative to 210 countries and territories worldwide, Israel ranked 47th in number of deaths per million citizens and 25th in number of confirmed cases per million citizens. In addition, the number of tests for coronavirus in Israel was 13,577 per million citizens [30].

Israel's management of the COVID-19 crisis differs from that of other countries because Israel is in the midst of a deep constitutional crisis, which is affecting its management of the health crisis. The spread of COVID-19 has created a multidimensional crisis in Israel. Internally, the epidemic is endangering public health, undermining economic and social resilience, challenging effective governance, and even providing cover for processes that could potentially harm democratic values [31].

Currently, the crisis in Israel is being managed by an interim government led by an interim prime minister operating under three criminal indictments after three election cycles in which there was no clear victor. Israel's parliament had stopped functioning normally [32]. Israeli Prime Minister Benjamin Netanyahu made critical decisions during the COVID-19 crisis, such as shutting down the court system and using surveillance to monitor patients, without bringing his decisions to the cabinet [32]. Some critics argue [33] that some of Prime Minister Netanyahu's decisions were motivated by personal interests (such as closing the courts and thereby postponing his trial) and by political considerations (including the decision to delay the lockdown of centers of contagion, such as ultra-Orthodox communities) [33-35].

In addition to the prime minister, the COVID-19 crisis in Israel is being managed by the director general of the Ministry of Health, who is an economist, and the head of Public Health Services at the Ministry of Health [36]. Most decisions are made by a small team established by Prime Minister Netanyahu [37]. Health Minister Yaakov Litzman (who is not a medical

professional) was involved in the decision-making process only at the beginning of the crisis; he refused to take measures against the ultra-Orthodox community, which is his constituency. During the crisis, hospital directors, physicians, and scientists criticized its management [37-39] and called for the replacement of the director general of the Ministry of Health and of Health Minister Yaakov Litzman by health care professionals [40,41]. There was also criticism of the shortages of testing kits [42] and protective equipment for medical workers [43] as well as of how conflicting information was communicated to the public.

In light of the unique confluence in Israel of the health crisis and the constitutional crisis, this study seeks to examine the response of the Israeli public to the Israeli government's emergency instructions regarding the pandemic in terms of correlations between overall risk perception and crisis management, overall risk perception and economic threat perception, crisis management and compliance with behavioral guidelines, and crisis management and economic threat perception. We also made comparisons between crisis management and spokesperson credibility and between crisis management and the credibility of information sources.

Methods

Sampling and Data Collection

The sample was planned using a Qualtrics XM online survey (Qualtrics Survey Software) that enabled rapid and effective distribution of an online questionnaire to our research population. The questionnaire is provided in Multimedia Appendix 1. We used the self-selection online survey method of nonprobability sampling [44] to recruit participants through social network posts asking the general public (people aged ≥18 years) to answer the survey. The rationale for using this sampling method is that the general public in Israel, including the researchers, was under movement restrictions at the time of the study; therefore, distribution of the questionnaire on social networks was more rapid and accessible.

The survey was distributed to the public using three main social media platforms: Facebook, WhatsApp, and Instagram. In the first stage, intensive sampling was accomplished through social networks and social media platforms. In the second stage, snowball sampling [45] was performed to reach broader circles in the Jewish and Arab communities. In the third stage, after a summary meeting and evaluation of the breakdown of the sociodemographic variables, it emerged that the number of participants from the Arab community was higher than that from the Jewish community. Another effort focused on dissemination through diverse circles in the Jewish community, such as community forums, official community Facebook pages, and internal diffusion circles that expanded to broader circles.

A total of 1348 people participated in the survey using the Qualtrics XM platform. However, 292/1348 (27.65%) of the questionnaires were not fully completed or were filled out by participants younger than 18 years. Those questionnaires were taken out of the sample, leaving a total of 1056 eligible participants.

The study was approved by the Faculty of Social Welfare and Health Sciences Ethics Committee for research with human subjects at the University of Haifa (Approval No. 20/088).

Research Tools

A quantitative questionnaire was designed to test the following variables: risk perception, crisis management, compliance with directives imposed on the public (report of behavioral intentions), and information sources. The questionnaire was based on previous questionnaires culturally accommodated to populations in Israel; it also accommodated the characteristics of the COVID-19 crisis and the measures taken in its wake [46-50].

Credibility and Validity

Before the questionnaire was distributed, a content validation process was undertaken by performing a pilot study of 20 participants in a limited sampling of the researchers' narrow circles. The participants were asked to provide feedback on the wording of the questionnaire, the time needed to fill out the questionnaire, etc., and changes were made accordingly. The questions were written in Hebrew and translated into Arabic; subsequently, changes were made in the wording to culturally accommodate it to the specific research population (eg, explanations were provided for certain statements to focus on the participants and prevent information bias).

Questionnaire Structure and Variable Design

In the first part of the questionnaire (Multimedia Appendix 1), the participants were asked to fill out their demographic information. The second part included an index of questions about risk perception. Some questions focused on the participant's personal fear of contracting coronavirus, such as "How serious do you perceive COVID-19 to be?" The personal risk perception index was the average of 2 questions (Cronbach α=.76). Other questions focused on fears about different age groups, such as "To what extent do you think the following populations are at high risk of contracting COVID-19?" The overall risk perception index was the average of 17 items (Cronbach α=0.91).

The third part of the questionnaire included an index of 14 questions (statements) on a Likert scale from 1 (not at all) to 5 (very) about the participants' behavior according to the guidelines given to the Israeli public as a whole. For example: "During the COVID-19 crisis, to what extent do you think you can give up the following behaviors: handshaking, hugging, kissing, not attending social gatherings, etc.?" and "How hard is it for you to follow the guidelines against leaving home to the following destinations?" Compliance to behavioral guidelines was calculated as the average of all 14 items (Cronbach α=.83).

The fourth part was an index of questions about perceptions of the management of the crisis, such as "I think that the measures taken by the state so far to prevent the spread of COVID-19 have been…" or "I think the Prime Minister's Office and the Ministry of Health are communicating the COVID-19 crisis to the public in a way that is…"

Attitudes on the crisis management index were calculated as the average of 3 questions after recoding the values of each question into 3 levels of evaluation: 1.5 for poor management, 3 for moderately good management, and 4.5 for good management (3 items, Cronbach α=.60).

The fifth part of the questionnaire included a question about economic security: "Beyond the health threat that COVID-19 poses for the public, to what extent does it threaten your economic security?" (1: It is no threat at all to 5: It is a very major threat.)

The sixth part included questions about the credibility of the spokesperson and credibility of the source of information, such as "What do you think is the most credible source of information on COVID-19?" (where participants were asked to mark one information source out of a list of sources) or "Do you feel you are receiving fully transparent information from the Ministry of Health?" (1: Not at all to 5: I receive extremely transparent information.)

Analysis

A comparison of the personal risk perceptions between age groups was tested using an analysis of variance (ANOVA) model in which the dependent variable was risk perception and the independent variable was age (4 age groups). The specific differences among age groups were tested by post hoc comparisons using the Tukey honestly significant difference (HSD) test.

Correlation between overall risk perception and attitudes toward crisis management and correlation between overall risk perception and economic threat perception were tested using the Pearson correlation coefficient.

Correlation between attitudes toward crisis management and compliance with behavioral guidelines and correlation between attitudes toward crisis management and economic threat perception were tested using the Pearson correlation coefficient.

The relationship between the most credible spokesperson and attitudes toward crisis management was tested using an ANOVA model in which the dependent variable was the attitude toward crisis management and the independent variable was the most credible spokesperson (6 groups, excluding family physicians or other). The specific differences among the 6 groups were tested by post hoc comparison using the Tukey HSD test.

The relationship between the most credible source of information and attitudes toward crisis management was tested using an ANOVA model in which the dependent variable was the attitudes toward crisis management and the independent variable was the source of information (6 sources). The specific differences between credible sources of information and the attitudes toward crisis management among the 6 groups were tested by post hoc comparisons using the Tukey HSD Test.

Results

Participants

A total of 1056 eligible participants filled out the online questionnaire (Table 1). 219/1056 (20.74%) were men and 837/1056 (79.26%) were women. The ages of the participants ranged from 18-95 years, with a mean age of 38 years. Of the participants, 423/1050 (40.29%) were Jewish and 627 (59.71%) were Arab. 395/1028 (38.42%) participants were secular, 443 (43.09%) were traditional, and 190 (18.48%) were religious. The participants' level of education was as follows: 118/1056 (11.17%) secondary education, 102 (9.66%) postsecondary education, 413 (39.11%) BA, 331 (31.34%) MA, 49 (4.64%) PhD, and 43 (4.07%) other types of education.

Because the ethnicity distribution of the sample was not proportional to the general population distribution, a weighting index was calculated. After weighting the data, the ethnicity distribution was 19% Arab and 81% Jewish according to the distribution of people aged ≥18 years in the general population in Israel. To reduce biases, weighting was applied to the data on all statistical inferences (the sociodemographic details presented in Table 1 are based on raw data with no weighting applied).

Table 1. Sociodemographic characteristics of the survey participants (N=1056).

Sociodemographic category and characteristics	n (%)
Gender	
Male	219 (20.74)
Female	837 (79.26)
Age (years)	
18-30	356 (33.78)
31-45	414 (39.28)
46-64	224 (21.25)
>65	60 (5.69)
Education	
Secondary	118 (11.17)
Postsecondary	102 (9.66)
BA	413 (39.11)
MA	331 (31.34)
PhD	49 (4.64)
Other	43 (4.07)
Ethnicity	
Jewish	423 (40.29)
Arab	627 (59.71)
Religion	
Secular	395 (38.42)
Traditional	443 (43.09)
Religious	190 (18.48)

Risk Perceptions

A significant difference in the personal risk perceptions between age groups ($F_{3,1050}$=5.14; P=.002) was detected. The means and standard deviations of the personal risk indices for the 4 age groups are presented in Table 2.

A significant difference was found between participants aged 65 years and older and participants aged 18-30 years or 46-64 years. Participants aged 65 years and older perceived higher personal risk compared to participants aged 18-30 years (mean difference 0.33, 95% CI 0.04-0.61) and compared to those aged 46-64 years (mean difference 0.38, 95% CI 0.12-0.64). There was no significant difference between participants aged 65 years and older and participants aged 31-45 years in the perception of their personal risk (Table 3).

A significant positive correlation (Pearson) between overall risk perception and attitude toward crisis management was found (r=0.19, P<.001). As risk perception increases, the evaluation of the crisis management tends to increase as well.

A significant positive correlation was found between overall risk perception and economic threat perception (r=0.22, P<.001). As risk perception increases, the evaluation of the economic threat tends to increase as well, and vice versa; higher economic threat perception is associated with higher risk perception.

Table 2. Comparison of the personal risk perceptions between age groups using an ANOVA model (N=1054).

Age group (years)	Personal risk index, mean (SD)
18-30	2.76 (0.55)
31-45	2.84 (0.67)
46-64	2.70 (0.74)
≥65	3.08 (0.94)

Table 3. Results of the Tukey HSD test for differences between personal risk perception and age group (N=1054).

Compared age groups (years)	Difference between means (95% CI)
≥65 and 31-45	0.25 (–0.01 to 0.50)
≥65 and 18-30	0.33 (0.04 to 0.61)[a]
≥65 and 46-64	0.38 (0.12 to 0.64)[a]
31-45 and 18-30	0.08 (–0.13 to 0.29)
31-45 and 46-64	0.14 (–0.04 to 0.32)
18-30 and 46-64	0.05 (–0.15 to 0.27)

[a]Statistically significant at α=.05.

Crisis Management

A significant positive correlation was found between attitudes toward crisis management and compliance with behavioral guidelines (r=0.15, *P*<.001). Higher evaluation of crisis management was associated with higher compliance.

A significant negative correlation was found between attitudes toward crisis management and economic threat perception (r=–0.15, *P*<.001). Higher economic threat perception was associated with lower evaluation of crisis management.

Spokesperson Credibility

A significant difference ($F_{5,981}$=43.16; *P*<.001) between participants who attributed the most credibility to different spokespersons and their attitudes toward crisis management was detected. The means and standard deviations of the attitudes toward crisis management for the six most credible spokespersons (n=987) are presented in Table 4.

Participants for whom the prime minister was the most credible spokesperson evaluated the crisis management significantly higher than all other groups (Table 5). Significantly lower evaluation of the crisis management was expressed by participants for whom infectious disease specialists were the most credible spokespersons compared to those who considered the director general of the Ministry of Health, the head of Public Health Services, or the Minister of Health to be most credible. Participants for whom journalists were the most credible spokespersons evaluated the crisis management significantly lower than those who believed that the director general of the Ministry of Health or the head of Public Health Services was the most credible.

Table 4. The relationships between the six most credible spokespersons and the attitudes toward crisis management using an ANOVA model (n=987).

Most credible spokesperson	Mean (SD)
Israeli Prime Minister	3.84 (0.44)
Director general of the Ministry of Health	3.58 (0.49)
Head of Public Health Services	3.49 (0.45)
Israeli Minister of Health	3.33 (0.34)
Infectious disease specialists	2.92 (0.68)
Journalists	2.87 (0.45)

Table 5. Results of the Tukey HSD test for differences between the most credible spokesperson and the attitudes toward crisis management (n=987).

Spokesperson comparison	Difference between means (95% CI)
Prime Minister	
Director general of the Ministry of Health	0.26 (0.03 to 0.48)[a]
Head of Public Health Services	0.35 (0.09 to 0.61)[a]
Minister of Health	0.51 (0.11 to 0.90)[a]
Infectious disease specialists	0.92 (0.70 to 1.14)[a]
Journalists	0.97 (0.48 to 1.45)[a]
Director general of the Ministry of Health	
Head of Public Health Services	0.09 (–0.13 to 0.31)
Minister of Health	0.25 (–0.12 to 0.62)
Infectious disease specialists	0.66 (0.49 to 0.83)[a]
Journalists	0.71 (0.24 to 1.19)[a]
Head of Public Health Services	
Minister of Health	0.15 (–0.24 to 0.55)
Infectious disease specialists	0.57 (0.35 to 0.78)[a]
Journalists	0.62 (0.13 to 1.11)[a]
Minister of Health	
Infectious disease specialists	0.41 (0.05 to 0.78)[a]
Journalists	0.46 (–0.11 to 1.04)
Infectious disease specialists	
Journalists	0.05 (–0.42 to 0.52)

[a]Statistically significant at α=.05.

Source of Information Credibility

A significant difference ($F_{5,1036}$=18.15; P<.001) was detected between participants who attributed the most credibility to different information sources and their attitudes toward crisis management. The means and standard deviations of the attitudes toward crisis management for the six most credible information sources are presented in Table 6.

Participants for whom the Ministry of Health website was the most credible source of information evaluated the crisis management higher than all other groups (Table 7).

Participants for whom scientific articles were the most credible source of information evaluated the crisis management lower than those who believed the WHO and Centers for Disease Control and Prevention (CDC) websites or health maintenance organization (HMO)/hospital websites and health care workers were the most credible.

Table 6. The relationships between the most credible information sources and the attitudes toward crisis management using an ANOVA model (N=1042).

Most credible information source	Mean (SD)
Ministry of Health website	3.61 (0.46)
WHO[a]/CDC[b] websites	3.23 (0.61)
HMO[c]/hospital websites and health care workers	3.19 (0.64)
Google/social networks	3.17 (0.68)
Media (television/newspapers)	3.07 (0.58)
Scientific articles	2.87 (0.77)

[a]WHO: World Health Organization.

[b]CDC: Centers for Disease Control and Prevention.

[c]HMO: health maintenance organization.

Table 7. Results of the Tukey HSD test for difference between most credible source of information and the attitudes toward crisis management (N=1042).

Source of information	Difference between means (95% CI)
Ministry of Health website	
WHO[a]/CDC[b] websites	0.38 (0.20-0.56)[c]
HMO[d]/hospital websites and health care workers	0.42 (0.20-0.65)[c]
Google/social networks	0.44 (0.02 to 0.87)[c]
Media (television/newspapers)	0.55 (0.17 to 0.93)[c]
Scientific articles	0.74 (0.49 to 0.99)[c]
WHO/CDC websites	
HMO/hospital websites and health care workers	0.04 (–0.18 to 0.26)
Google/social networks	0.06 (–0.37 to 0.49)
Media (television/newspapers)	0.16 (–0.21 to 0.54)
Scientific articles	0.36 (0.11 to 0.61)[c]
HMO/hospital websites and health care workers	
Google/social networks	0.02 (–0.42 to 0.47)
Media (television/newspapers)	0.13 (–0.27 to 0.52)
Scientific articles	0.32 (0.04 to 0.60)[c]
Google/social networks	
Media (television/newspapers)	0.10 (–0.44 to 0.64)
Scientific articles	0.29 (–0.17 to 0.76)
Media (television/newspapers)	
Scientific articles	0.19 (–0.23 to 0.61)

[a]WHO: World Health Organization.

[b]CDC: Centers for Disease Control and Prevention.

[c]Statistically significant at α=.05.

[d]HMO: health maintenance organization.

Discussion

Principal Findings

This study was conducted during March 2020 in Israel. We sought to examine public perceptions of risk concerning COVID-19 and public assessment of policymakers' management of the crisis. The findings indicate that participants aged 65 years and older have a higher risk perception of contracting COVID-19 than the younger age groups. This finding is consistent with scientific facts indicating that older people are at highest risk due to the severity of the illness and the fatality rate [51]. The older age group indicated that its fears are science-based and were not false concerns. Similarly, the younger age groups were less afraid, which is consistent with their relative risk.

The findings of this study also indicate that the greater the participant's personal risk perception, the better they evaluated the crisis management, and vice versa. A possible explanation of this finding is that people who are very concerned about COVID-19 are at such a high level of fear that they view any action taken by policymakers to confront and combat the virus as reasonable. Furthermore, the actions taken by Israel, such as requiring people returning from abroad to enter quarantine since the beginning of the crisis before there were any fatalities in Israel; grounding of flights; cancellation of public events and gatherings; surveillance and phone tracking of patients; and curfew on the entire public were perceived by people who are very afraid of the disease as appropriate and not excessive. To reinforce this interpretation, it was found in studies that higher levels of perceived susceptibility are associated with greater intention to change behavior in the manner recommended in the fear appeal message; also, a higher level of perceived susceptibility is a strong determinant of intentions and behavior, even in the face of weak arguments [52]. On the other hand, people with lower perceived susceptibility evaluated the crisis management as less good, possibly for the opposite reason: they view the draconian measures taken by Israel, including an arbitrary curfew on the entire population (including subpopulations that were not at risk) and a curfew on geographical areas where contagion was low, as excessive and disproportionate. These measures have drawn public criticism [38,53].

Another finding of the study is a positive significant correlation between overall risk perception and economic threat perception. As risk perceptions increase, the evaluation of the economic threat also tends to increase, and vice versa. This finding indicates that the health crisis caused by the COVID-19 pandemic had far-reaching consequences for the global, national and personal economy; therefore, the participants were afraid not only of the health threat but also the inherent economic threat. It was also found that higher economic threat perception was associated with lower evaluation of crisis management.

It is likely that people who perceive a high personal economic threat feel that the government is not managing the crisis well if it is allowing their economic resilience to be harmed. The feeling that the crisis management is causing fatal harm to the Israeli economy also arises from conversations on social networks and the Israeli media, where it has been argued that the high economic and political price that Israel is paying is even more dangerous than COVID-19 [54].

Another key finding from this study was a positive significant correlation between attitudes toward crisis management and compliance with behavioral guidelines. Studies indicate that public trust in government institutions and leaders is considered essential in any country that seeks to impose its authority on the public and maintain order. High evaluation and trust of the functioning of authorities affects the behavior of the public [11,12]. In the context of this study, the participants' high evaluation of the crisis management in Israel affected the public's high compliance with the guidelines during the pandemic. This finding reinforces the importance of trust in the health care system, especially during a crisis such as the COVID-19 pandemic, when the public is asked to change its routine behaviors and habits.

Furthermore, the findings of this study indicate that participants for whom the prime minister was the most credible spokesperson evaluated the crisis management as significantly better than all other groups. The crisis management was evaluated as significantly worse by participants who perceived infectious disease specialists to be the most credible spokespersons compared to those who perceived the most credible spokespersons to be the director general of the Ministry of Health, the head of Public Health Services, or the Minister of Health. These findings indicate the importance of spokespersons during epidemic crises [55].

Selecting appropriate spokespersons to communicate with the public during and after a health crisis is a strategic decision that can have far-reaching results [56-58]. The spokesperson is perceived as the representative of the establishment managing the crisis [55,59,60]. The higher the credibility of the spokesperson, the greater the chance the audience will be open to receiving the messages and complying with the guidelines.

Participants who viewed Prime Minister Netanyahu as the most credible spokesperson evaluated the crisis management as good because he performs two functions: manager and spokesperson of the crisis. Throughout the crisis, Prime Minister Netanyahu appeared at dozens of press conferences and delivered the guidelines to the public himself. Conversely, participants who viewed infectious disease specialists as the most credible spokespersons provided a lower evaluation of the crisis management. The apparent reason is that in Israel, the crisis was managed over the entire period since the COVID-19 crisis began by a very narrow and centralized team [37]. This team was harshly criticized by certain elements in the health care system and the general public. The criticism focused on the insufficient number of public health and medical experts on the team. Due to this criticism, during the crisis, hospital directors and physicians called for the director general of the Ministry of Health, who is an economist, and the health minister, who is not a health professional, to be dismissed and replaced with professionals [40,61]. In epidemic/pandemic crises in the age of new media, it is important for both spokespersons and information sources to be perceived by the public as credible [62-65]. The findings of this study indicate that those who perceived the health ministry to be the most credible information source also perceived the crisis management to be the most favorable, contrary to those who perceived academic articles to be the most credible information sources and perceived the crisis management as less favorable.

It is likely that the participants who perceived the Ministry of Health website (ie, the website that represents the body managing the crisis) as most credible will also perceive the crisis management to be good. Thus, they are exposed to information on the Ministry of Health website, which supports its management decisions with findings and testimonies; this exposure apparently affects their view of the crisis management as optimal. Conversely, people who read academic articles and are not only exposed to informative materials provided to them by the Ministry of Health are likely to be highly literate; therefore, it is likely that they are exposed to other materials and findings that are not consistent with the Ministry of Health guidelines.

Experts have argued that the Israeli Ministry of Health guidelines are contradictory. At the beginning of the COVID-19 crisis, the Israeli Ministry of Health claimed that there was no need for masks, although they were already being used in other countries according to recommendations and prior knowledge [66]; however, later, the guideline was changed to require the public to use masks. In another example of conflicting information, due to the shortage of personal protective equipment (PPE) for its employees, the Ministry of Health initially issued a statement in mid-March stating that health care workers do not need to wear PPE regularly but only in certain situations [67]. Following the Ministry of Health statement, senior physicians from across the country claimed that the Ministry of Health statement was an excuse to cover up the inadequacy of the Israeli health care system. According to the last State Comptroller and Ombudsman of Israel report [68], the PPE shortage is only one example of this inadequacy.

Also, it was found in the literature that vaccine-hesitant groups who show skepticism toward the establishment are exposed to academic articles and do not rely only on government information [49]. Follow-up studies can examine the association between the phenomenon of hesitancy and how hesitant groups perceive the management of the COVID-19 crisis.

Limitations

The limitations of this study are that it is not a representative study. This study used nonprobability sampling procedures and measuring. Despite the nonprobability sampling, the sample included a high total number of participants. Secondly, since the research was conducted during the COVID-19 crisis and it was important to examine the public's positions regarding crisis management, we decided to distribute the survey online on social networks to reach a broad circle of people in a short time. Furthermore, during the COVID-19 crisis, the public was required to maintain social distancing; therefore, an online survey was the most suitable tool. However, the sociodemograpic statistics presented suggest that a diverse sample was reached based on sociodemographic variables. Since the ethnicity distribution of the sample was not proportional to the general population distribution, a weighting index was calculated. The ethnicity distribution after weighting the data was 19% Arab and 81% Jewish according to the distribution of the general population aged ≥18 years in Israel. To reduce biases, weighting was applied to the data on all statistical inferences.

Conclusions

This study suggests that it is critical to establish public trust in decision makers. The higher the public trust and evaluation of crisis management, the more the public will comply with guidelines. It was also found that the crisis management and information cannot be approached in the same way for the overall public. Decision makers must address and communicate the risks differently to different subpopulations that have different risk perceptions and different levels of health literacy. Furthermore, unlike other epidemic crises, the COVID-19 crisis has widespread economic and social consequences; therefore, it is impossible to communicate and focus only on the health risk without communicating the economic and social risks as well.

Acknowledgments

The authors wish to thank the study participants.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The online survey questionnaire.
[DOCX File , 33 KB - jmir_v22i5e19370_app1.docx]

References

1. World Health Organization. Geneva, Switzerland; 2020 Jan 30. Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV) URL: https://tinyurl.com/rjdtx2k [accessed 2020-04-14]
2. World Health Organization. Geneva, Switzerland Coronavirus disease 2019 (COVID-19) situation reports URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports [accessed 2020-05-13]
3. World Health Organization. Geneva, Switzerland Coronavirus disease (COVID-19) advice for the public URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public [accessed 2020-05-13]
4. Betsch C, Wieler L, Habersaat K. Monitoring behavioural insights related to COVID-19. Lancet 2020 Apr;395(10232):1255-1256 [FREE Full text] [doi: 10.1016/s0140-6736(20)30729-7]
5. Clements J. medRxiv. Knowledge and behaviors toward COVID-19 among U.S. residents during the early days of the pandemic URL: https://www.medrxiv.org/content/medrxiv/early/2020/04/02/2020.03.31.20048967.full.pdf [accessed 2020-04-14]
6. Barry JM. Pandemics: avoiding the mistakes of 1918. Nature 2009 May 21;459(7245):324-325 [FREE Full text] [doi: 10.1038/459324a] [Medline: 19458695]
7. Ratzan SC, Moritsugu KP. Ebola crisis--communication chaos we can avoid. J Health Commun 2014;19(11):1213-1215. [doi: 10.1080/10810730.2014.977680] [Medline: 25356719]
8. Gesser-Edelsburg A, Mordini E, James JJ, Greco D, Green MS. Risk Communication Recommendations and Implementation During Emerging Infectious Diseases: A Case Study of the 2009 H1N1 Influenza Pandemic. Disaster Med Public Health Prep 2014 Apr 15:1-12. [doi: 10.1017/dmp.2014.27] [Medline: 24735822]
9. Gesser-Edelsburg A, Shir-Raz Y. Risk Communication and Infectious Diseases in an Age of Digital Media. London and New York: Routledge; Nov 10, 2016.
10. Bensman J. The aesthetics and politics of footnoting. Int J Polit Cult Soc 1988 Mar;1(3):443-470. [doi: 10.1007/bf01385430]
11. Devos T, Spini D, Schwartz S. Conflicts among human values and trust in institutions. Br J Soc Psychol 2002 Dec;41(Pt 4):481-494. [doi: 10.1348/014466602321149849] [Medline: 12593749]

12. Tang L. The influences of patient's trust in medical service and attitude towards health policy on patient's overall satisfaction with medical service and sub satisfaction in China. BMC Public Health 2011 Jun 15;11:472 [FREE Full text] [doi: 10.1186/1471-2458-11-472] [Medline: 21676228]
13. Meyer S, Ward P, Coveney J, Rogers W. Trust in the health system: An analysis and extension of the social theories of Giddens and Luhmann. Health Sociol Rev 2014 Dec 17;17(2):177-186. [doi: 10.5172/hesr.451.17.2.177]
14. Gesser-Edelsburg A, Walter N, Shir-Raz Y, Green MS. Voluntary or Mandatory? The Valence Framing Effect of Attitudes Regarding HPV Vaccination. J Health Commun 2015;20(11):1287-1293. [doi: 10.1080/10810730.2015.1018642] [Medline: 26132725]
15. Ozawa S, Paina L, Qiu M. Exploring pathways for building trust in vaccination and strengthening health system resilience. BMC Health Serv Res 2016 Nov 15;16(Suppl 7):639 [FREE Full text] [doi: 10.1186/s12913-016-1867-7] [Medline: 28185595]
16. Centers for Disease Control and Prevention. Atlanta, GA; 2020 Apr 03. Interim infection prevention and control recommendations for patients with suspected or confirmed Coronavirus disease 2019 (COVID-19) in healthcare settings URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/infection-control-recommendations.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Finfection-control%2Fcontrol-recommendations.html [accessed 2020-04-14]
17. European Centre for Disease Prevention and Control. relief.int. 2020 Mar 31. Infection prevention and control for COVID-19 in healthcare settings: Third update URL: https://reliefweb.int/sites/reliefweb.int/files/resources/Infection-prevention-control-for-the-care-of-patients-with-2019-nCoV-healthcare-settings_update-31-March-2020.pdf [accessed 2020-04-14]
18. Alsheikh R. Ynet. 2020 Mar 25. Road loss in the treatment of the corona pandemic. Article in Hebrew URL: https://www.ynet.co.il/articles/0,7340,L-5701260,00.html [accessed 2020-04-14]
19. Scripps Research Institute. ScienceDaily. 2020 Mar 17. COVID-19 coronavirus epidemic has a natural origin URL: https://www.sciencedaily.com/releases/2020/03/200317175442.htm [accessed 2020-04-14]
20. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. Nat Med 2020 Apr;26(4):450-452 [FREE Full text] [doi: 10.1038/s41591-020-0820-9] [Medline: 32284615]
21. Gesser-Edelsburg A, Shir-Raz Y. Science vs. fear: the Ebola quarantine debate as a case study that reveals how the public perceives risk. J Risk Res 2015 Nov 05:1-23. [doi: 10.1080/13669877.2015.1100659]
22. Gesser-Edelsburg A, Shir-Raz Y, Green MS. Why do parents who usually vaccinate their children hesitate or refuse? General good vs. individual risk. J Risk Res 2014 Dec 12;19(4):405-424. [doi: 10.1080/13669877.2014.983947]
23. Gebrekidan S. The New York Times. 2020 Mar 12. The World Has a Plan to Fight Coronavirus. Most Countries Are Not Using it URL: https://www.nytimes.com/2020/03/12/world/coronavirus-world-health-organization.html [accessed 2020-04-14]
24. Landler M. The New York Times. 2020 Mar 11. A fumbled global response to the virus in a leadership void URL: https://www.nytimes.com/2020/03/11/world/europe/coronavirus-leadership-trump.html [accessed 2020-04-14]
25. Dudden A, Marks A. The Guardian. 2020 Mar 20. South Korea took rapid, intrusive measures against Covid-19 – and they worked URL: https://www.theguardian.com/commentisfree/2020/mar/20/south-korea-rapid-intrusive-measures-covid-19 [accessed 2020-04-14]
26. Rapoza K. Forbes. 2020 Mar 26. How Singapore and South Korea deal with Coronavirus quarantine measures URL: https://www.forbes.com/sites/kenrapoza/2020/03/26/how-singapore-and-south-korea-deal-with-coronavirus-quarantine-measures/#7e4398751921 [accessed 2020-04-14]
27. Fisher M, Sang-Hun C. The New York Times. 2020 Mar 23. How South Korea flattened the curve URL: https://www.nytimes.com/2020/03/23/world/asia/coronavirus-south-korea-flatten-curve.html [accessed 2020-04-14]
28. Hjelmgaard K. USA Today. 2020 Mar 17. These countries are doing the best and worst jobs fighting coronavirus URL: https://www.usatoday.com/story/news/world/2020/03/17/coronavirus-how-countries-across-globe-responding-covid-19/5065867002/ [accessed 2020-04-14]
29. Druckman Y. Ynet. 2020 Mar 20. From the recommendation not to go to China to the ban on leaving the house: That's how the Corona crisis unfolded. Article in Hebrew URL: https://www.ynet.co.il/articles/0,7340,L-5700913,00.html [accessed 2020-04-14]
30. Worldometer. 2020 Apr 12. COVID-19 Coronavirus pandemic URL: https://www.worldometers.info/coronavirus/ [accessed 2020-04-13]
31. Dekel U. The Institute for National Security Studies. 2020 Mar 26. The system-wide campaign to battle the Corona crisis in Israel URL: https://www.inss.org.il/publication/coronavirus-special-publication/ [accessed 2020-04-14]
32. Schneider T. Globes. 2020 Mar 15. Interpretation: Israel's information policy raises more questions than answers. Article in Hebrew URL: https://www.globes.co.il/news/article.aspx?did=1001321748#utm_source=Apps&utm_medium=SharedTo [accessed 2020-04-14]
33. Padan C, Elran M, Levi E, Hadad S, Ben AA. The Institute for National Security Studies. 2020 Mar 20. The Corona Epidemic: Systemic Challenges for Israel URL: https://www.inss.org.il/publication/coronavirus-and-the-challenges-for-israel/ [accessed 2020-04-14]

34. Gal I. Ynet. 2020 Mar 29. Final: Corona does not spread through the air - how does it spread? Article in Hebrew URL: https://www.ynet.co.il/health/article/HkEwO11RIL [accessed 2020-04-14]
35. Harel A. Haaretz. 2020 Apr 07. Passover curfew is a sign it's all political in Israel, even the Coronavirus pandemic URL: https://www.haaretz.com/israel-news/.premium-even-in-a-pandemic-it-s-all-political-1.8746312 [accessed 2020-04-14]
36. Verthaim D. Globes. 2020 Mar 18. Opinion: The betrayal of the media - sponsored by the Corona, the news editions abused their role. Article in Hebrew URL: https://www.globes.co.il/news/article.aspx?did=1001322424 [accessed 2020-04-14]
37. Efrati I, Kashti O. Haaretz. 2020 Mar 23. Netanyahu is coordinating the fight in Corona, and senior health officials remain outside. Article in Hebrew URL: https://www.haaretz.co.il/health/corona/.premium-1.8700253 [accessed 2020-04-14]
38. Arlosoroff M. TheMarker Magazine. 2020 Apr 20. Managemental crises, not a health one: The default that prevents a return to routine. Article in Hebrew URL: https://www.themarker.com/coronavirus/.premium-1.8761487?utm_source=App_Share&utm_medium=iOS_Native [accessed 2020-04-14]
39. Ashkenazi S. Globes. 2020 Mar 30. Delivered clear messages based on data, not fear. Article in Hebrew URL: https://www.globes.co.il/news/article.aspx?did=1001323860 [accessed 2020-04-14]
40. Ashkenazi S. Globes. 2020 Mar 29. Doctors v. Litzman: Require Netanyahu to replace Health Minister. Article in Hebrew URL: https://www.globes.co.il/news/article.aspx?did=1001323705 [accessed 2020-04-14]
41. Libsker A. Calcalist. 2020 Mar 12. Corona is braking. Humanity will not go extinct. Article in Hebrew URL: https://newmedia.calcalist.co.il/magazine-12-03-20/m01.html?fbclid=IwAR0u-BR_SauOwu3eWyT0ORbMy_m7Eg2Dv9zvBDmXLLPsGwM2v44Ss-kCzbM [accessed 2020-04-14]
42. Arlosoroff M. TheMarker Magazine. 2020 Mar 20. That is not how the fight the epidemic: The failure of the Corona tests will cost us 90 billion NIS. Article in Hebrew URL: https://www.themarker.com/news/health/.premium-1.8684362?utm_source=App_Share&utm_medium=iOS_Native [accessed 2020-04-14]
43. Blau U. Haaretz. 2020 Mar 20. Official documents reveal: Lack of protective and emergency equipment, authorities in the equipping race URL: https://www.haaretz.co.il/health/corona/.premium-1.8697312?utm_source=App_Share&utm_medium=iOS_Native [accessed 2020-04-14]
44. Vehovar V, Toepoel V, Steinmetz S. Chapter 22: Non-probability Sampling. In: Wolf C, Joye D, Smith TW, Fu YC, editors. The SAGE Handbook of Survey Methodology. Newbury Park, CA: SAGE Publications Ltd; 2016:329-345.
45. Ulin P, Robinson E, Tolley E. Qualitative Methods in Public Health: A Field Guide for Applied Research. San Francisco, CA: Jossey-Bass; 2004.
46. Gesser-Edelsburg A, Abed Elhadi Shahbari N, Cohen R, Mir Halavi A, Hijazi R, Paz-Yaakobovitch G, et al. Differences in Perceptions of Health Information Between the Public and Health Care Professionals: Nonprobability Sampling Questionnaire Survey. J Med Internet Res 2019 Jul 03;21(7):e14105 [FREE Full text] [doi: 10.2196/14105] [Medline: 31271145]
47. Gesser-Edelsburg A, Walter N, Green MS. Health care workers--part of the system or part of the public? Ambivalent risk perception in health care workers. Am J Infect Control 2014 Aug;42(8):829-833. [doi: 10.1016/j.ajic.2014.04.012] [Medline: 24939516]
48. Gesser-Edelsburg A, Zemach M. From a fiasco to the Supertanker grand finale: Israeli Prime Minister Netanyahu's crisis communication during the Carmel disaster. J Risk Res 2012 Sep;15(8):967-989. [doi: 10.1080/13669877.2012.686052]
49. Gesser-Edelsburg A, Walter N, Shir-Raz Y, Sassoni Bar-Lev O, Rosenblat S. The behind-the-scenes activity of parental decision-making discourse regarding childhood vaccination. Am J Infect Control 2017 Mar 01;45(3):267-271. [doi: 10.1016/j.ajic.2016.10.009] [Medline: 27856072]
50. Gesser-Edelsburg A, Shalayeva S. Internet as a Source of Long-Term and Real-Time Professional, Psychological, and Nutritional Treatment: A Qualitative Case Study Among Former Israeli Soviet Union Immigrants. J Med Internet Res 2017 Feb 03;19(2):e33 [FREE Full text] [doi: 10.2196/jmir.7130] [Medline: 28159729]
51. Centers for Disease Control and Prevention. Atlanta, GA Coronavirus disease 2019 (COVID-19) - Older adults URL: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/older-adults.html [accessed 2020-04-14]
52. de Hoog N, Stroebe W, de Wit JBF. The impact of fear appeals on processing and acceptance of action recommendations. Pers Soc Psychol Bull 2005 Jan;31(1):24-33. [doi: 10.1177/0146167204271321] [Medline: 15574659]
53. Qimron U, Gavish U. Ynet. 2020 Apr 13. Maybe the vaccine is already with us? Article in Hebrew URL: https://www.ynet.co.il/articles/0,7340,L-5713835,00.html [accessed 2020-04-14]
54. Rolnik G. TheMarker Magazine. 2020 Apr 01. The political virus is more dangerous than the Corona URL: https://www.themarker.com/magazine/.premium-1.8730744 [accessed 2020-04-14]
55. Kim S, Johnson E. Comprising or compromising credibility? Use of spokesperson quotations in news releases issued by major health agencies. PR Journal 2008;2(4):1-18 [FREE Full text]
56. Bauman D. Evaluating Ethical Approaches to Crisis Leadership: Insights from Unintentional Harm Research. J Bus Ethics 2010 Jul 21;98(2):281-295. [doi: 10.1007/s10551-010-0549-3]
57. James EH, Wooten LP. Leadership as (Un)usual: How to Display Competence in Times of Crisis. Organ Dyn 2005 Jan;34(2):141-152. [doi: 10.1016/j.orgdyn.2005.03.005]
58. Marsen S. Navigating Crisis: The Role of Communication in Organizational Crisis. Int J Bus Commun 2019 Oct 21;57(2):163-175. [doi: 10.1177/2329488419882981]

59. Bennett P, Calman K, Curtis S, Smith D. Risk Communication and Public Health, Second Edition. Oxford: Oxford University Press; 2010.
60. Heath R, O'Hair H. Handbook of Risk and Crisis Communication. New York, NY: Routledge; 2009.
61. Yanko A. Calcalist. 2020 Mar 20. Department heads to Netanyahu and Gantz: A professional health minister needs to be appointed urgently instead of Litzman. Article in Hebrew URL: https://www.calcalist.co.il/local/articles/0,7340,L-3804573,00.html [accessed 2020-04-14]
62. Gesser-Edelsburg A, Shir-Raz Y, Walter N, Mordini E, Dimitriou D, James JJ, et al. The Public Sphere in Emerging Infectious Disease Communication: Recipient or Active and Vocal Partner? Disaster Med Public Health Prep 2015 Aug;9(4):447-458. [doi: 10.1017/dmp.2015.31] [Medline: 26186579]
63. Boyd DM, Ellison NB. Social Network Sites: Definition, History, and Scholarship. J Comput Mediat Commun 2007;13(1):210-230 [FREE Full text] [doi: 10.1111/j.1083-6101.2007.00393.x]
64. Castells M. The New Public Sphere: Global Civil Society, Communication Networks, and Global Governance. Ann Am Acad Pol Soc Sci 2008 Mar;616(1):78-93. [doi: 10.1177/0002716207311877]
65. Gesser-Edelsburg A, Walter N, Shir-Raz Y. The "New Public" and the "Good Ol' Press": Evaluating Online News Sources During the 2013 Polio Outbreak in Israel. Health Commun 2017 Feb;32(2):169-179. [doi: 10.1080/10410236.2015.1110224] [Medline: 27191949]
66. MacIntyre CR, Chughtai AA. Facemasks for the prevention of infection in healthcare and community settings. BMJ 2015 Apr 09;350:h694. [doi: 10.1136/bmj.h694] [Medline: 25858901]
67. Gal I. Ynet. 2020 Mar 21. Ministry of Health guidance to physicians and pharmacist: No routine PPE is necessary. Article in Hebrew URL: https://www.ynet.co.il/articles/0,7340,L-5699125,00.html [accessed 2020-04-14]
68. Jerusalem: The State Comptroller and Ombudsman of Israel. The hospitalization distress press release URL: https://www.mevaker.gov.il/he/publication/Articles/Pages/2015.02.04-Hospital.aspx?AspxAutoDetectCookieSupport=1 [accessed 2020-04-14]

Abbreviations

ANOVA: analysis of variance
CDC: Centers for Disease Control and Prevention
COVID-19: coronavirus disease
HMO: Health Maintenance Organization
HSD: honestly significant difference
PPE: personal protective equipment
SARS: severe acute respiratory syndrome
WHO: World Health Organization

Edited by G Eysenbach; submitted 19.04.20; peer-reviewed by P Banik, C García, J McCool; comments to author 07.05.20; revised version received 07.05.20; accepted 11.05.20; published 15.05.20.

Please cite as:
Gesser-Edelsburg A, Cohen R, Hijazi R, Abed Elhadi Shahbari N
Analysis of Public Perception of the Israeli Government's Early Emergency Instructions Regarding COVID-19: Online Survey Study
J Med Internet Res 2020;22(5):e19370
URL: http://www.jmir.org/2020/5/e19370/
doi:10.2196/19370
PMID:32392172

Original Paper

Mining the Characteristics of COVID-19 Patients in China: Analysis of Social Media Posts

Chunmei Huang[1*], MD; Xinjie Xu[2*], BMed; Yuyang Cai[3*], PhD; Qinmin Ge[2*], PhD; Guangwang Zeng[1,4*], BMed; Xiaopan Li[5,6], PhD; Weide Zhang[7], MD; Chen Ji[8], PhD; Ling Yang[1], PhD

[1]Department of Geriatrics, Xinhua Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China

[2]Department of Emergency, Xinhua Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China

[3]School of Public Health, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[4]The Health Center of Nansheng Town, Wuzhishan, China

[5]Center for Disease Control and Prevention, Pudong New Area, Shanghai, China

[6]Pudong Institute of Preventive Medicine, Pudong New Area, Fudan University, Shanghai, China

[7]Big Data and Artificial Intelligence Center, Zhongshan Hospital, Fudan University, Shanghai, China

[8]Warwick Clinical Trials Unit, Warwick Medical School, Coventry, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Ling Yang, PhD
Department of Geriatrics
Xinhua Hospital
Shanghai Jiaotong University School of Medicine
1665 Kongjiang Road
Shanghai
China
Phone: 86 13651608005
Email: yangling01@xinhuamed.com.cn

Abstract

Background: In December 2019, pneumonia cases of unknown origin were reported in Wuhan City, Hubei Province, China. Identified as the coronavirus disease (COVID-19), the number of cases grew rapidly by human-to-human transmission in Wuhan. Social media, especially Sina Weibo (a major Chinese microblogging social media site), has become an important platform for the public to obtain information and seek help.

Objective: This study aims to analyze the characteristics of suspected or laboratory-confirmed COVID-19 patients who asked for help on Sina Weibo.

Methods: We conducted data mining on Sina Weibo and extracted the data of 485 patients who presented with clinical symptoms and imaging descriptions of suspected or laboratory-confirmed cases of COVID-19. In total, 9878 posts seeking help on Sina Weibo from February 3 to 20, 2020 were analyzed. We used a descriptive research methodology to describe the distribution and other epidemiological characteristics of patients with suspected or laboratory-confirmed SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) infection. The distance between patients' home and the nearest designated hospital was calculated using the geographic information system ArcGIS.

Results: All patients included in this study who sought help on Sina Weibo lived in Wuhan, with a median age of 63.0 years (IQR 55.0-71.0). Fever (408/485, 84.12%) was the most common symptom. Ground-glass opacity (237/314, 75.48%) was the most common pattern on chest computed tomography; 39.67% (167/421) of families had suspected and/or laboratory-confirmed family members; 36.58% (154/421) of families had 1 or 2 suspected and/or laboratory-confirmed members; and 70.52% (232/329) of patients needed to rely on their relatives for help. The median time from illness onset to real-time reverse transcription-polymerase chain reaction (RT-PCR) testing was 8 days (IQR 5.0-10.0), and the median time from illness onset to online help was 10 days (IQR 6.0-12.0). Of 481 patients, 32.22% (n=155) lived more than 3 kilometers away from the nearest designated hospital.

Conclusions: Our findings show that patients seeking help on Sina Weibo lived in Wuhan and most were elderly. Most patients had fever symptoms, and ground-glass opacities were noted in chest computed tomography. The onset of the disease was characterized by family clustering and most families lived far from the designated hospital. Therefore, we recommend the

following: (1) the most stringent centralized medical observation measures should be taken to avoid transmission in family clusters; and (2) social media can help these patients get early attention during Wuhan's lockdown. These findings can help the government and the health department identify high-risk patients and accelerate emergency responses following public demands for help.

(J Med Internet Res 2020;22(5):e19087) doi:10.2196/19087

KEYWORDS

SARS-CoV-2; COVID-19; coronavirus disease; social media; Sina Weibo; help

Introduction

Background

In December 2019, pneumonia cases of unknown origin were reported in Wuhan City, Hubei Province, China. The illness was identified and officially named as coronavirus disease 2019 (COVID-19), which is caused by a novel viral strain called severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [1-3] and resembles severe acute respiratory syndrome coronavirus (SARS-CoV) [4]. Since the outbreak, COVID-19 has spread rapidly. Person-to-person transmission in hospital and family settings had occurred due to close contact [5,6]. On January 23, 2020, Wuhan shut down public transportation and was placed under lockdown, and residents were not allowed to leave the city. As of February 20, 2020, the accumulative number of laboratory-confirmed patients in Wuhan was 45,346. The health care system was further overburdened as patients with mild symptoms sought hospitalization instead of self-isolation, mainly due to the anxiety and panic instigated by the epidemic [7]. After failing to be admitted to a hospital, patients sought help on Sina Weibo, a Chinese microblogging site similar to Twitter that allows people to communicate and share information instantly [8]. Social media has become an important channel for promoting risk communication during the crisis [9,10] and can be used to measure public attention given to public health emergencies [11], such as H7N9 [12-14], Ebola [9,15-19], Zika virus [10,20,21], Middle East respiratory syndrome (MERS-CoV) [22], and Dengue fever [23].

Since the COVID-19 outbreak, social media, especially Sina Weibo, has become an important platform for the public to obtain epidemic-related information quickly and effectively. According to the official outbreak data released by Sina Weibo on February 26, 2020, 51.2 million users cumulatively posted 350 million pieces of epidemic-related content. Online readership of epidemic-related topics reached 754.5 billion. Sina Weibo established a communication channel that allowed the government to effectively listen and respond to public opinion quickly. Here, by collecting data from Sina Weibo from February 3 to 20, 2020, we aim to analyze the characteristics of suspected or laboratory-confirmed patients with the SARS-CoV-2 infection.

Objective

In this study, we describe the characteristics of suspected or laboratory-confirmed patients with the SARS-CoV-2 infection, the distribution of patients throughout Wuhan, and the relationship between helpers (eg, relative, friend, spouse, sibling) and patients. Social media was used to obtain timely access to public demand so that the government and the health department could identify high-risk patients and take measures to help these patients.

Methods

Overview

Sina Weibo launched a platform to provide online help channels for patients infected with SARS-CoV-2. From February 3 to 20, 2020, we obtained 9878 posts by using the keyword 肺炎患者求助 (COVID-19 pneumonia patients seeking help) from Sina Weibo through its application programming interface (API). Python (Python Software Foundation) was used to implement a rule-based screening and classification method on the PyCharm platform. We used the collected posts as a training set, including related posts and unrelated posts. Based on the post-for-help rules formulated by Sina Weibo, we considered the post text, as well as keywords pertaining to name, age, home address, time of illness, and description of illness as a related post; otherwise, it was deemed an irrelevant post. We excluded 6922 irrelevant posts that only described opinions and feelings about help seeking related to COVID-19 and initially collected 2956 related posts that contained mentions of clinical symptoms and/or imaging descriptions. Then, we manually screened out and excluded posts. We excluded 1679 reposted posts, 556 posts with a significant amount of missing valid clinical data, 195 nonpneumonia patient posts, and 41 patient posts with non-Wuhan home addresses. Finally, we selected 485 patient posts that presented clinical symptoms and imaging descriptions (Figures 1 and 2). The number of patient posts on Sina Weibo has been declining because these patients have actively deleted posts upon hospital admission.

We collected clinical symptoms, chest computed tomography (CT) findings (the chest CT was only summarized for those who provided a clinical report), days from illness onset to online help, days from illness onset to RT-PCR testing, RT-PCR test results, the relationship between helpers and patients, and home address details from Sina Weibo's records. We performed a study on the clinical characteristics of suspected or laboratory-confirmed patients with the SARS-CoV-2 infection seeking help on Sina Weibo. Suspected cases were identified as having fever or respiratory symptoms such as shortness of breath, cough, productive sputum, or chest pain. A laboratory-confirmed case with SARS-CoV-2 infection was defined as a positive result to high throughput sequencing or real-time reverse transcription-polymerase chain reaction (RT-PCR) assay of throat swabs and sputum [2].

We also used a descriptive research methodology to analyze the distribution of patients throughout Wuhan and the relationship between helpers and patients. The distance from patients' home to the nearest designated hospital was calculated using the geographic information system ArcGIS. The data used in the current study is publicly accessible on Sina Weibo and readers can obtain the raw data online [24]. We have effectively protected the privacy of subjects and strictly adhered to the principle of confidentiality in terms of information collection, storage and transmission, and information use and deletion. The study was approved by the Shanghai Jiaotong University Xinhua Hospital Ethics Committee and was carried out in accordance with the Declaration of Helsinki. We have made an application for exemption from informed consent and obtained approval.

Figure 1. An example of a patient with coronavirus disease (COVID-19) seeking help on Sina Weibo. RT-PCR: reverse transcription-polymerase chain reaction.

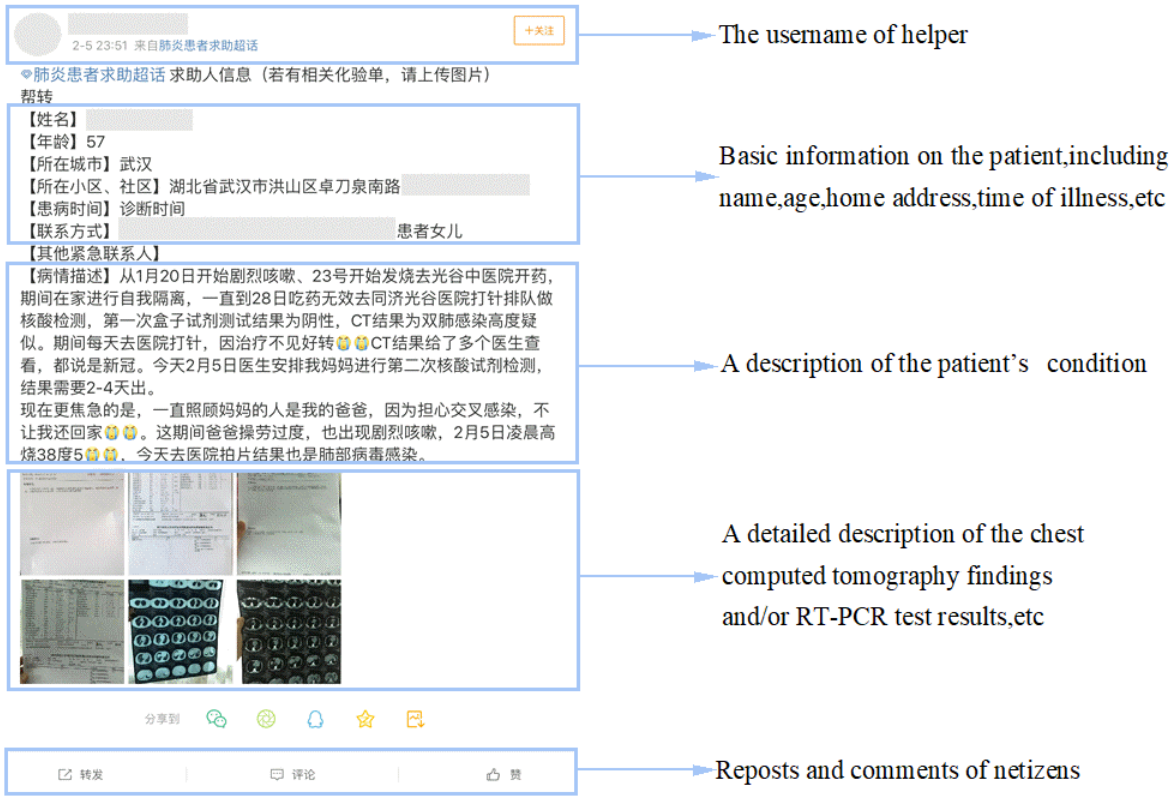
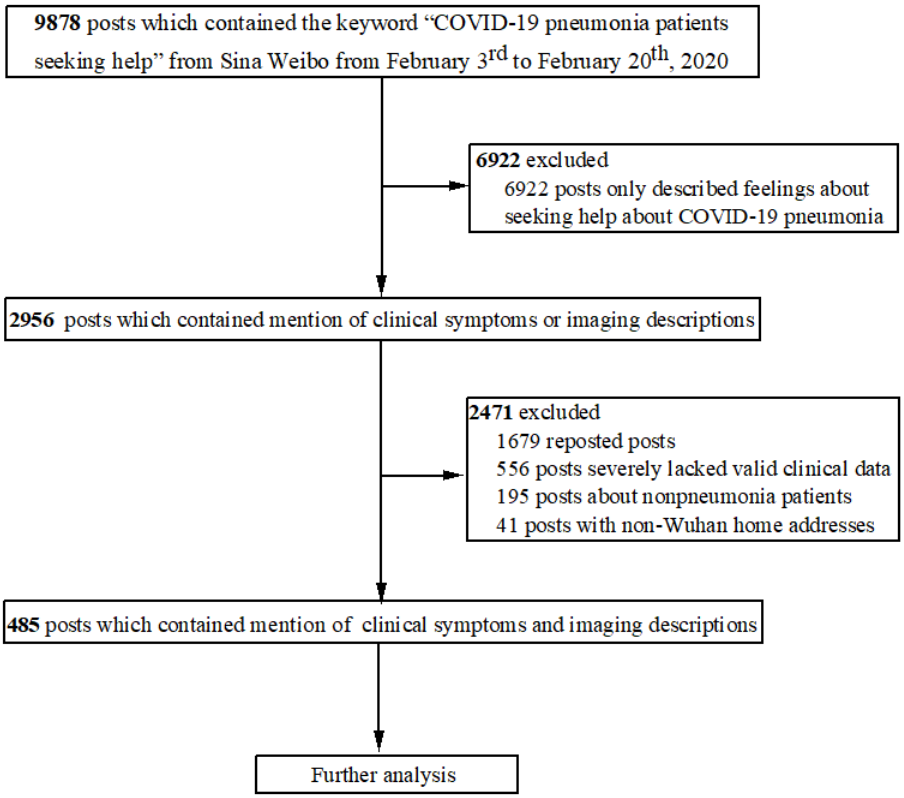


Figure 2. Study flow diagram. COVID-19: coronavirus disease.



Statistical Analysis

Continuous variables were expressed as median (IQR) when appropriate. Categorical variables were summarized as counts and percentages in each category. Analysis was conducted using SPSS, version 19.0 (IBM). We used ArcGIS, version 10.2.2, to plot the numbers of patients seeking help on a map.

Results

Demographic and Clinical Characteristics

We selected 485 patients with suspected or laboratory-confirmed SARS-CoV-2 infection with at least clinical symptoms and imaging descriptions from Sina Weibo. The demographic and clinical characteristics were shown in Table 1. The median age was 63.0 years (IQR 55.0 to 71.0), 0.21% (1/470) of patients were below 15 years of age, and 50.10% (243/485) were female. Fever (408/485, 84.12%) was the most common symptom. Other symptoms reported by patients included fatigue (224/485, 46.19%), shortness of breath (261/485, 53.81%), nausea or vomiting (81/485, 16.70%) and diarrhea (61/485, 12.58%). In total, 23.09% (112/485) of patients had at least one underlying disorder (eg, hypertension, chronic obstructive pulmonary disease, etc). All patients underwent chest CT. Of these patients, 35.26% (171/485) reported lung infection on the chest CT but did not provide their clinical reports. In the remaining 64.74% (314/485) of patients, the most common pattern on chest CT was ground-glass opacity (237/314, 75.48%) and bilateral patchy shadowing (191/314, 60.83%). The median time from illness onset to RT-PCR testing was 8.0 days (IQR 5.0-10.0), and the median time from illness onset to online help was 10.0 days (IQR 6.0-12.0). RT-PCR testing was performed in 52.16% (253/485) of patients; 68.38% (173/253) were positive, 1.98% (5/253) were suspected cases, and 10.67% (27/253) were negative.

The 485 patients came from 421 families, and 39.67% (167/421) of these families had at least one family member with a laboratory-confirmed and/or suspected diagnosis of SARS-CoV-2; 11.40% (48/421) of families had one laboratory-confirmed family member only. Families with one confirmed case accounted for 9.50% (40/421); two, three, and four confirmed members accounted for 1.19% (5/421), 0.48% (2/421), and 0.24% (1/421), respectively. A suspected diagnosis occurred in 30.64% (129/421) of families; a family with one suspected member accounted for 21.14% (89/421), two suspected members accounted for 7.60% (32/421), three suspected members accounted for 1.43% (6/421), and four suspected members accounted for 0.48% (2/421) (Figure 3).

Table 1. Clinical characteristics of suspected or laboratory-confirmed patients with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection (N=485).

Characteristic	Value
Age (years), median (IQR)	63.0 (55.0-71.0)
Age groups (n=470), n (%)	
0-14 years	1 (0.21)
15-49 years	74 (15.74)
50-64 years	178 (37.87)
≥65 years	217 (46.17)
Sex (female; n=485), n (%)	243 (50.10)
Respiratory symptoms, n (%)	
Fever (temperature ≥37.3°C)	408 (84.12)
Cough	190 (39.18)
Fatigue	224 (46.19)
Shortness of breath	261 (53.81)
Nausea or vomiting	81 (16.70)
Diarrhea	61 (12.58)
Coexisting disorders (n=485), n (%)	
Any	112 (23.09)
Chronic obstructive pulmonary disease	10 (2.06)
Diabetes	43 (8.87)
Hypertension	55 (11.34)
Coronary heart disease	38 (7.84)
Cerebrovascular diseases	12 (2.47)
Cancer[a]	7 (1.44)
Chronic renal diseases	7 (1.44)
Immunodeficiency	2 (0.41)
Hepatitis B infection[b]	3 (0.62)
Radiologic findings: abnormalities on chest CT [c] (n=314), n (%)	
Ground-glass opacity	237 (75.48)
Local patchy shadowing	20 (6.37)
Bilateral patchy shadowing	191 (60.83)
Interstitial abnormalities	6 (1.91)
Days from illness onset to online help, median (range)	10 (6-12)
Days from illness onset to RT-PCR[d] testing, median (range)	8 (5-10)
Underwent RT-PCR testing (n=253), n (%)	253 (52.16)
Positive	173 (68.38)
No result	48 (18.97)
Suspect	5 (1.98)
Negative	2 (10.67)

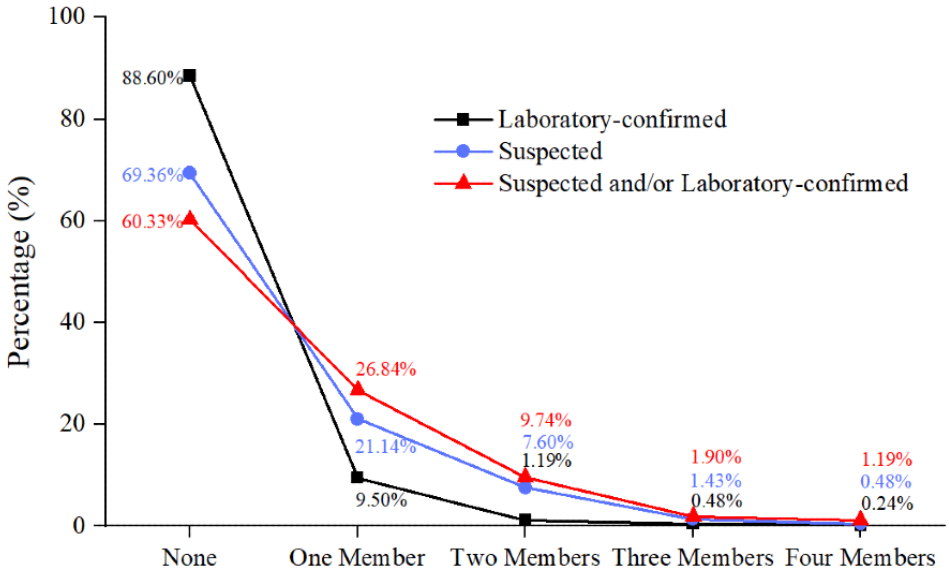
[a]Cancer referred to any malignancy.

[b]Hepatitis B infection denotes a positive test for hepatitis B surface antigen, with or without elevated alanine or aspartate aminotransferase levels.

[c]CT: computed tomography.

[d]RT-PCR: reverse transcription-polymerase chain reaction.

Figure 3. The distribution of family clusters.



The Distribution of Patients Throughout Wuhan and the Distance Between Helpers and Hospitals

All patients were located in Wuhan, but more patients lived in the central districts (Hongshan, Jiang'an, Wuchang, Hanyang, and Qiaokou) compared to outskirt districts (Figure 4). We further analyzed the distance between patients and the nearest designated hospital. Among these patients, four had missing home address information. We found that 25.57% (123/481) were within 1 kilometer of the nearest designated hospital, 24.74% (119/481) lived within 1-2 kilometers, 17.46% (84/481) lived within 2-3 kilometers, and 32.22% (155/481) lived more than 3 kilometers away (Table 2 and Figure 5).

Figure 4. The distribution of patients throughout Wuhan.

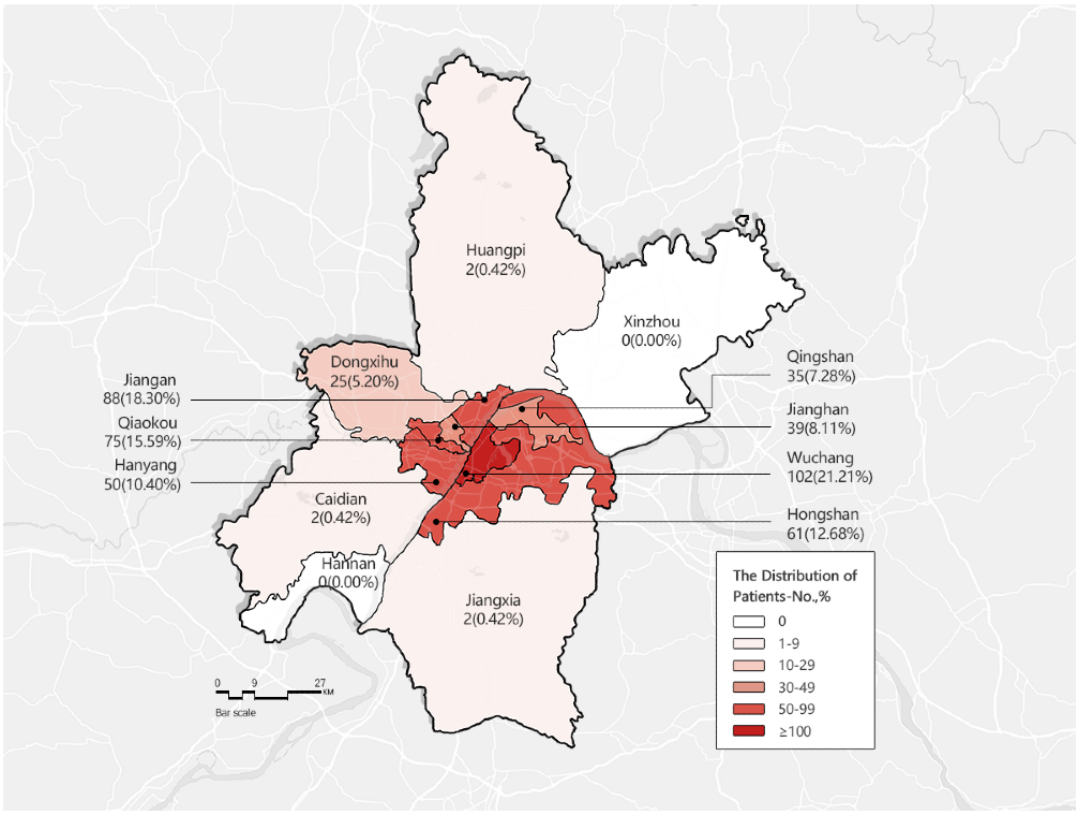
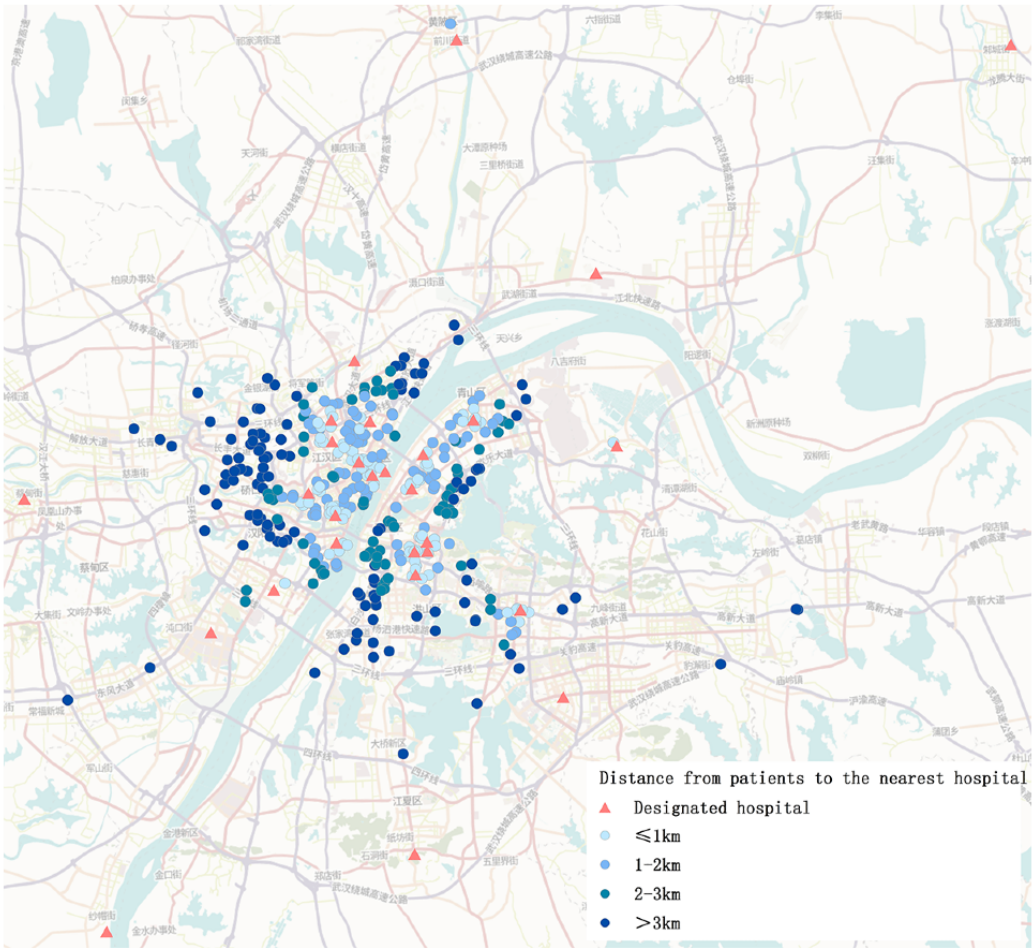


Table 2. The distance between patients and the nearest designated hospital, as well as the relationship between helpers and patients.

Variable	Count, n (%)
Distance (n=481)	
≤1km	123 (25.57)
1-2 km	119 (24.74)
2-3 km	84 (17.46)
≥3km	155 (32.22)
Relationship between helper and patient (n=329)	
Relative	232 (70.52)
Friend	38 (11.55)
Patient themselves	34 (10.33)
Spouse	14 (4.26)
Sibling	11 (3.34)

Figure 5. The distance between patients and their nearest designated hospital.



The Relationship Between Helpers and Patients

We explored the relationship between helpers and patients. During data collection, 156 helpers stated that they were family members of patients, but they did not specify their relationship. The remaining 70.52% (232/329) of helpers were the patients' relatives; 11.55% (38/329) were friends, 4.26% (14/329) were their spouses, 3.34% (11/329) were siblings, and 10.33% (34/329) were the patients themselves (Table 2).

Discussion

Principal Findings

This study has shown that patients seeking help on Sina Weibo lived in Wuhan and most of them were elderly. Our statistical analysis of the age of patients seeking help on Sina Weibo demonstrated that patients on Sina Weibo were older—the proportion of patients who were ≥65 years was as high as

46.17%. Zhong et al [3] reported that only a small proportion (15.1%) of 1099 laboratory-confirmed COVID-19 patients were aged ≥65 years. On the other hand, our study has found that the highest incidence was among adults over 50 years of age [25].

Additionally, 23.09% of patients had at least one underlying disorder. Fever was the dominant symptom whereas gastrointestinal symptoms were rare. Ground-glass opacity was the most common pattern on chest CT. Among all laboratory-confirmed COVID-19 patients, the most common pattern on chest CT were ground-glass opacity (56.4%) [3]. Our study has shown that the median time from illness onset to RT-PCR testing was 8 days, and the median time from illness onset to online help was 10 days. A recent study showed that the mean time from onset to hospital admission in 44 patients in Wuhan, with onset before January 1, was 12.5 days; in 189 patients with onset from January 1 to 11, the mean time was 9.1 days [5].

Person-to-person transmission of COVID-19 in hospital and family settings has been increasing [26-29]. Family clustering played an important part in increasing the number of COVID-19 cases [30]. Our study provided further evidence of human-to-human transmission, although 60.33% of families had no clustered onset, indicating that home isolation may be effective for patients. However, 39.67% of families had suspected and/or laboratory-confirmed cases among family members. In addition, 36.58% of families had 1 or 2 suspected and/or laboratory-confirmed family members. This is also in line with the finding that patients, on average, transmit the infection to 2.2 other people [5]. Therefore, home isolation might lead to the risk of COVID-19 outbreaks in family clusters [31]. This means that it is crucial to strictly isolate patients and trace and quarantine contacts as early as possible [32,33]. The most stringent centralized medical observation measures should be taken as soon as possible to avoid outbreaks in family clusters due to home isolation [31], such as a modular hospital to treat patients with mild illness [34].

Our research also found that the number of patients in the Wuchang, Jiang'an, Qiaokou, Hongshan, and Hanyang districts was greater than in other districts. Figure 4 shows a central agglomeration of patients; this may be consistent with the outbreak of the epidemic in the Huanan Seafood Wholesale Market in the Jianghan district, which was thought to be the initial infection site from an animal source in China [35] or it may be related to the developed economy, convenient transportation, and the population density in the city center. Therefore, close contact with family members and actual population movements from the outbreak source were risk factors for the spread of SARS-CoV-2.

In total, 32.22% (155/481) of patients lived more than 3 kilometers away from their nearest designated hospital. According to Baidu maps, adults can walk 4 kilometers in 1 hour. Considering that the patients in this study were older and their health condition may have slowed them down even more, we estimate that patients could walk 3 kilometers in a 1-hour period. Hence, this indicates that a patient would need to walk more than 1 hour to see a doctor since public transportation was suspended at the time. This may be one of the reasons why patients wanted to be admitted to a hospital. In addition, on February 5th, the Wuhan municipal health commission designated 28 hospitals for the treatment of laboratory-confirmed patients with the SARS-CoV-2 infection. The empty bed rate of hospitals within the city was only 3.6%. Thus, patients could not be hospitalized for the various reasons above. This also reflected an insufficiency of medical resources during the initial outbreak [36].

We also explored the relationship between helpers and patients. Judging from the content of Sina Weibo posts asking for help, “Mom” and “Dad” were high-frequency words; 70.52% (232/329) of helpers were the patients' relatives, indicating that the publishers of the help information were mostly the children of the elderly. Unfamiliarity with new technology may have hindered elderly people from seeking assistance from the outside world.

With the rapid and effective dissemination of help information, since February 5th, the People's Daily has launched an all-media operation to provide online help channels for patients with the SARS-CoV-2 infection. The government implemented a policy to maximize hospital admissions, which led to a rapid decrease in the number of people seeking help on Sina Weibo on February 6th and remained at low levels since February 8th, indicating that the needs of the public had been met. This also means that it is important to establish new and effective communication mechanisms for the dissemination of important factual information in a timely manner. Through this epidemic, we can see that medical resources are insufficiently allocated. There are substantial regional disparities in health care resource availability and accessibility in China [37]. The rapid increase in the number of patients during the initial outbreak led to a relative shortage of medical resources, which may threaten people with poor self-help capabilities such as the elderly. The government and health departments should pay attention to the elderly population during the outbreak. Social media can be used to understand public demand and aid the government in formulating accurate countermeasures following public demands for help. Although social media can establish effective communication channels, this technology may require a certain threshold, so the government should continue to increase the availability and accessibility of the network to better respond to public health emergencies.

Limitations

Our study has some limitations. First, given that our data was collected from a social media platform, the description of patients’ symptoms and laboratory information were likely to be incomplete. Second, the urgent timeline for data extraction and the subjective judgment of the collectors might undermine the data quality to a certain extent. Finally, we learned that most of these patients have been admitted to the hospital with government help and many patients remain in the hospital, so we did not compare the 28-day rate for the composite endpoint.

Conclusions

In summary, our study found that the distance between patients and hospitals and the closure of public transportation further increased the difficulty of hospitalization for the elderly. We

recommend the application of centralized medical observations to avoid the spread of COVID-19 through family clusters. In a public health emergency, making full use of available social media platforms can establish effective, factual communication channels and shorten admission times, helping patients get early attention during the Wuhan lockdown. These findings can help the government and health departments pay attention to the elderly population during the outbreak and accelerate emergency responses following public demands for help.

Acknowledgments

This study was supported by the Zhejiang University Special Scientific Research Fund for COVID-19 prevention and control (2020XGZX065); the National Natural Science Foundation of China (71432006); the National Social Science Fund of China (grant number 17BSH056); and the Shanghai Jiao Tong University think tank research project (ZKYJ-20200114). The funders had no role in the study design, data collection, data analysis and interpretation, writing of the report, or the decision to submit the paper for publication. The Authors were also supported by the Shanghai Jiaotong University special scientific research fund for "Technology Promotion Project" in 2019 (ZT201903) and the Shanghai science and Technology Fund (18441905200).

Authors' Contributions

CH, XX, YC, QG, and GZ are joint first authors. LY, QG, and YC are joint co-correspondence authors. LY and YC obtained funding. CH, XX, PL, YC, QG, and LY participated in study design; CH, XX, WZ, and GZ collected the data; CH, XX, GZ, WZ, and PL performed data analysis; CH and JC drafted the manuscript; CH, XX, PL, YC, QG, and LY were responsible for study conception. All authors provided critical review of the manuscript and approved the final draft for publication. QG, YC, and LY contributed to the interpretation of the results and critical revision of the manuscript for important intellectual content and approved the final version of the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Hu D, Zhu C, Ai L, He T, Wang Y, Ye F, et al. Genomic characterization and infectivity of a novel SARS-like coronavirus in Chinese bats. Emerg Microbes Infect 2018 Sep 12;7(1):154 [FREE Full text] [doi: 10.1038/s41426-018-0155-5] [Medline: 30209269]
2. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet 2020 Feb 15;395(10223):497-506 [FREE Full text] [doi: 10.1016/S0140-6736(20)30183-5] [Medline: 31986264]
3. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. 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] [Medline: 32109013]
4. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, China Novel Coronavirus Investigating Research Team. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med 2020 Feb 20;382(8):727-733 [FREE Full text] [doi: 10.1056/NEJMoa2001017] [Medline: 31978945]
5. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. N Engl J Med 2020 Mar 26;382(13):1199-1207 [FREE Full text] [doi: 10.1056/NEJMoa2001316] [Medline: 31995857]
6. Huang R, Xia J, Chen Y, Shan C, Wu C. A family cluster of SARS-CoV-2 infection involving 11 patients in Nanjing, China. The Lancet Infectious Diseases 2020 May 28;20(5):534-535. [doi: 10.1016/S1473-3099(20)30147-X] [Medline: 32119823]
7. Jakovljevic M, Bjedov S, Jaksic N, Jakovljevic I. COVID-19 Pandemia and Public and Global Mental Health from the Perspective of Global Health Securit. Psychiatr Danub 2020;32(1):6-14 [FREE Full text] [doi: 10.24869/psyd.2020.6] [Medline: 32303023]
8. Wang Z, Yu G, Tian X, Tang J, Yan X. A Study of Users with Suicidal Ideation on Sina Weibo. Telemed J E Health 2018 Sep;24(9):702-709. [doi: 10.1089/tmj.2017.0189] [Medline: 29293069]
9. Househ M. Communicating Ebola through social media and electronic news media outlets: A cross-sectional study. Health Informatics J 2016 Sep 26;22(3):470-478. [doi: 10.1177/1460458214568037] [Medline: 25656678]
10. Gui X, Wang Y, Kou Y, Reynolds TL, Chen Y, Mei Q, et al. Understanding the Patterns of Health Information Dissemination on Social Media during the Zika Outbreak. AMIA Annu Symp Proc 2017;2017:820-829 [FREE Full text] [Medline: 29854148]
11. Zhao Y, Cheng S, Yu X, Xu H. Chinese Public's Attention to the COVID-19 Epidemic on Social Media: Observational Descriptive Study. J Med Internet Res 2020 May 04;22(5):e18825 [FREE Full text] [doi: 10.2196/18825] [Medline: 32314976]

12. Gu H, Chen B, Zhu H, Jiang T, Wang X, Chen L, et al. Importance of Internet surveillance in public health emergency control and prevention: evidence from a digital epidemiologic study during avian influenza A H7N9 outbreaks. J Med Internet Res 2014;16(1):e20 [FREE Full text] [doi: 10.2196/jmir.2911] [Medline: 24440770]
13. Chen Y, Zhang Y, Xu Z, Wang X, Lu J, Hu W. Avian Influenza A (H7N9) and related Internet search query data in China. Sci Rep 2019 Jul 18;9(1):10434 [FREE Full text] [doi: 10.1038/s41598-019-46898-y] [Medline: 31320681]
14. Xie T, Yang Z, Yang S, Wu N, Li L. Correlation between reported human infection with avian influenza A H7N9 virus and cyber user awareness: what can we learn from digital epidemiology? Int J Infect Dis 2014 May;22:1-3 [FREE Full text] [doi: 10.1016/j.ijid.2013.11.013] [Medline: 24576842]
15. Fung IC, Fu K, Chan C, Chan BSB, Cheung C, Abraham T, et al. Social Media's Initial Reaction to Information and Misinformation on Ebola, August 2014: Facts and Rumors. Public Health Rep 2016;131(3):461-473 [FREE Full text] [doi: 10.1177/003335491613100312] [Medline: 27252566]
16. Fung IC, Duke CH, Finch KC, Snook KR, Tseng P, Hernandez AC, et al. Ebola virus disease and social media: A systematic review. Am J Infect Control 2016 Dec 01;44(12):1660-1671. [doi: 10.1016/j.ajic.2016.05.011] [Medline: 27425009]
17. Wong R, Harris JK. Geospatial Distribution of Local Health Department Tweets and Online Searches About Ebola During the 2014 Ebola Outbreak. Disaster Med Public Health Prep 2018 Jun 24;12(3):287-290. [doi: 10.1017/dmp.2017.69] [Medline: 28835312]
18. Feng S, Hossain L, Crawford JW, Bossomaier T. Quantifying Network Dynamics and Information Flow Across Chinese Social Media During the African Ebola Outbreak. Disaster Med Public Health Prep 2018 Feb 01;12(1):26-37. [doi: 10.1017/dmp.2017.29] [Medline: 28760166]
19. Seltzer E, Jean N, Kramer-Golinkoff E, Asch D, Merchant R. The content of social media's shared images about Ebola: a retrospective study. Public Health 2015 Sep;129(9):1273-1277. [doi: 10.1016/j.puhe.2015.07.025] [Medline: 26285825]
20. Seltzer EK, Horst-Martz E, Lu M, Merchant RM. Public sentiment and discourse about Zika virus on Instagram. Public Health 2017 Sep;150:170-175. [doi: 10.1016/j.puhe.2017.07.015] [Medline: 28806618]
21. Fu K, Liang H, Saroha N, Tse ZTH, Ip P, Fung IC. How people react to Zika virus outbreaks on Twitter? A computational content analysis. Am J Infect Control 2016 Dec 01;44(12):1700-1702. [doi: 10.1016/j.ajic.2016.04.253] [Medline: 27566874]
22. Fung IC, Fu K, Ying Y, Schaible B, Hao Y, Chan C, et al. Chinese social media reaction to the MERS-CoV and avian influenza A(H7N9) outbreaks. Infect Dis Poverty 2013;2(1):31 [FREE Full text] [doi: 10.1186/2049-9957-2-31] [Medline: 24359669]
23. Guo P, Zhang Q, Chen Y, Xiao J, He J, Zhang Y, et al. An ensemble forecast model of dengue in Guangzhou, China using climate and social media surveillance data. Sci Total Environ 2019 Jan 10;647:752-762 [FREE Full text] [doi: 10.1016/j.scitotenv.2018.08.044] [Medline: 30092532]
24. Xu X. Mining the characteristics of COVID-19 patients from social media in China (Raw data).xlsx (Dataset). Figshare 2020:e [FREE Full text]
25. Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. [The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China]. Zhonghua Liu Xing Bing Xue Za Zhi 2020 Feb 17;41(2):145-151. [doi: 10.3760/cma.j.issn.0254-6450.2020.02.003] [Medline: 32064853]
26. Chan JF, Yuan S, Kok K, To KK, Chu H, Yang J, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. The Lancet 2020 Feb 15;395(10223):514-523 [FREE Full text] [doi: 10.1016/S0140-6736(20)30154-9] [Medline: 31986261]
27. Phan LT, Nguyen TV, Luong QC, Nguyen TV, Nguyen HT, Le HQ, et al. Importation and Human-to-Human Transmission of a Novel Coronavirus in Vietnam. N Engl J Med 2020 Feb 27;382(9):872-874 [FREE Full text] [doi: 10.1056/NEJMc2001272] [Medline: 31991079]
28. Rothe C, Schunk M, Sothmann P, Bretzel G, Froeschl G, Wallrauch C, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. N Engl J Med 2020 Mar 05;382(10):970-971. [doi: 10.1056/NEJMc2001468] [Medline: 32003551]
29. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet 2020 Feb 29;395(10225):689-697. [doi: 10.1016/S0140-6736(20)30260-9] [Medline: 32014114]
30. Guan Q, Liu M, Zhuang YJ, Yuan Y, Wang SS, Li J, et al. [Epidemiological investigation of a family clustering of COVID-19]. Zhonghua Liu Xing Bing Xue Za Zhi 2020 Mar 09;41:629-633. [doi: 10.3760/cma.j.cn112338-20200223-00152] [Medline: 32149484]
31. Qiu YY, Wang SQ, Wang XL, Lu WX, Qiao D, Li JB, et al. [Epidemiological analysis on a family cluster of COVID-19]. Zhonghua Liu Xing Bing Xue Za Zhi 2020 Mar 05;41(4):506-509. [doi: 10.3760/cma.j.cn112338-20200221-00147] [Medline: 32133831]
32. Yang HY, Xu J, Li Y, Liang X, Jin YF, Chen SY, et al. [The preliminary analysis on the characteristics of the cluster for the Corona Virus Disease]. Zhonghua Liu Xing Bing Xue Za Zhi 2020 Mar 08;41:623-628. [doi: 10.3760/cma.j.cn112338-20200223-00153] [Medline: 32145716]
33. Tian S, Hu N, Lou J, Chen K, Kang X, Xiang Z, et al. Characteristics of COVID-19 infection in Beijing. J Infect 2020 Apr;80(4):401-406. [doi: 10.1016/j.jinf.2020.02.018] [Medline: 32112886]

34. Yusong L, Duanliang P, Jia Y, Dunyan C, Hongbing J, Siyuan Y, et al. Laboratory diagnostics within a modular hospital at the time of Coronavirus disease 2019 (COVID-19) in Wuhan. Clin Chem Lab Med 2020 Apr 02:e [FREE Full text] [doi: 10.1515/cclm-2020-0332] [Medline: 32242830]
35. Liu K, Fang Y, Deng Y, Liu W, Wang M, Ma J, et al. Clinical characteristics of novel coronavirus cases in tertiary hospitals in Hubei Province. Chin Med J (Engl) 2020 May 05;133(9):1025-1031 [FREE Full text] [doi: 10.1097/CM9.0000000000000744] [Medline: 32044814]
36. Ji Y, Ma Z, Peppelenbosch MP, Pan Q. Potential association between COVID-19 mortality and health-care resource availability. Lancet Glob Health 2020 Apr;8(4):e480 [FREE Full text] [doi: 10.1016/S2214-109X(20)30068-1] [Medline: 32109372]
37. Yu M, He S, Wu D, Zhu H, Webster C. Examining the Multi-Scalar Unevenness of High-Quality Healthcare Resources Distribution in China. Int J Environ Res Public Health 2019 Aug 07;16(16):e [FREE Full text] [doi: 10.3390/ijerph16162813] [Medline: 31394765]

Abbreviations

API: application programming interface
COVID-19: coronavirus disease
CT: computed tomography
MERS-CoV: Middle East respiratory syndrome
RT-PCR: reverse transcription-polymerase chain reaction
SARS-CoV: severe acute respiratory syndrome coronavirus
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2

Edited by G Eysenbach; submitted 05.04.20; peer-reviewed by A Benis, J Mueller, D Surian; comments to author 20.04.20; revised version received 03.05.20; accepted 12.05.20; published 17.05.20.

Please cite as:
Huang C, Xu X, Cai Y, Ge Q, Zeng G, Li X, Zhang W, Ji C, Yang L
Mining the Characteristics of COVID-19 Patients in China: Analysis of Social Media Posts
J Med Internet Res 2020;22(5):e19087
URL: http://www.jmir.org/2020/5/e19087/
doi:10.2196/19087
PMID:32401210

Original Paper

The Impact of Social Media on Panic During the COVID-19 Pandemic in Iraqi Kurdistan: Online Questionnaire Study

Araz Ramazan Ahmad[1,2*], MA, PhD; Hersh Rasool Murad[3*], MA, PhD

[1]Department of Administration, College of Humanities, University of Raparin, Ranya - Al Sulaimaniyah, Iraq

[2]Department of International Relations & Diplomacy, Faculty of Administrative Sciences and Economics, Tishk International University, Erbil, Iraq

[3]Department of Public Relations & Marketing, Technical College of Administration, Sulaimani Polytechnic University, Sulaimani, Iraq

[*]all authors contributed equally

Corresponding Author:
Araz Ramazan Ahmad, MA, PhD
Department of Administration
College of Humanities
University of Raparin
Main Road
Ranya - Al Sulaimaniyah, 46012
Iraq
Phone: 964 7701573122
Email: araz.ahmad85@uor.edu.krd

Abstract

Background: In the first few months of 2020, information and news reports about the coronavirus disease (COVID-19) were rapidly published and shared on social media and social networking sites. While the field of infodemiology has studied information patterns on the Web and in social media for at least 18 years, the COVID-19 pandemic has been referred to as the first social media *infodemic*. However, there is limited evidence about whether and how the social media infodemic has spread panic and affected the mental health of social media users.

Objective: The aim of this study is to determine how social media affects self-reported mental health and the spread of panic about COVID-19 in the Kurdistan Region of Iraq.

Methods: To carry out this study, an online questionnaire was prepared and conducted in Iraqi Kurdistan, and a total of 516 social media users were sampled. This study deployed a content analysis method for data analysis. Correspondingly, data were analyzed using SPSS software.

Results: Participants reported that social media has a significant impact on spreading fear and panic related to the COVID-19 outbreak in Iraqi Kurdistan, with a potential negative influence on people's mental health and psychological well-being. Facebook was the most used social media network for spreading panic about the COVID-19 outbreak in Iraq. We found a significant positive statistical correlation between self-reported social media use and the spread of panic related to COVID-19 (R=.8701). Our results showed that the majority of youths aged 18-35 years are facing psychological anxiety.

Conclusions: During lockdown, people are using social media platforms to gain information about COVID-19. The nature of the impact of social media panic among people varies depending on an individual's gender, age, and level of education. Social media has played a key role in spreading anxiety about the COVID-19 outbreak in Iraqi Kurdistan.

(J Med Internet Res 2020;22(5):e19556) doi:10.2196/19556

KEYWORDS

social media; COVID-19; infodemic; panic; mental health; fake news; misinformation; impact; Kurdistan region; Iraq

Introduction

Background

The coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered coronavirus [1]. Cases of COVID-19 first emerged in late December 2019, when a mysterious illness was reported in Wuhan, China. The cause of the disease was soon confirmed as a novel coronavirus, and the infection has since spread to many countries worldwide and has become a pandemic disease [2]. Several websites have published

information about COVID-19 and have given different instructions to their users about ways to prevent the spread of the virus, such as keeping a distance between themselves and others, using masks, and washing their hands [3]. Social media has become a source of disseminating information to the public. Many individuals will experience isolation during hospitalization or when quarantining at home [4]. Social media can be an efficient source of information and an effective means for staying abreast of the vast amount of medical knowledge [5].

COVID-19 Cases in Iraqi Kurdistan

Prior to the outbreak of COVID-19, people already relied on social media to gather information and news, and since the outbreak in December 2019, people in many countries have relied on social media to obtain information about the virus. In addition, people in Iraqi Kurdistan depend on social media. Internet use is strongly associated with behaviors related to health information; users write about their health on various social media platforms [6]. According to Kemp [7], there were 29.82 million internet users and 21 million social media users in Iraq in January 2020. Therefore, internet data, including data from social media platforms such as Twitter, have been used extensively in the past 20 years to study health patterns and better understand infectious disease outbreaks, a field known as infodemiology or (if used as surveillance tool) infoveillance [8].

At the time of writing, the global spread of COVID-19 is still a rapidly evolving situation. The Kurdistan Regional Government (KRG) created a webpage [9], which is regularly updated by the Ministry of Health, to communicate information about COVID-19 cases in Kurdistan. According to government statements, only the Ministry of Health or the World Health Organization (WHO) is qualified to confirm COVID-19 cases in Iraqi Kurdistan. Nevertheless, most people rely on social media and look for information on social media platforms instead of using the official KRG webpage.

According to statements from the KRG's Ministry of Health, as of April 10, 2020, the total number of confirmed cases is 324, including 3 deaths, 134 recovered patients, and 187 active cases [9]. The main objective of this early study is to investigate the relationship between using social media platforms and the spreading of panic during this COVID-19 outbreak.

Literature Review

The first study on social media during a pandemic dates back to the 2009 H1N1 pandemic, tracking the prevalence of misinformation (determined as 4.5%), terminology use ("H1N1" versus "swine flu"), public sentiments and fear, and relationships between case incidence and public concern [10]. Previous studies used the internet to collect data related to diseases, such as the search frequency of hand washing, hand sanitizer, and antiseptic topics [11]. The WHO declared that they are currently fighting not only an international epidemic but also a social media infodemic, with some media claiming that the coronavirus is the first true social media infodemic because it has accelerated information and misinformation worldwide and is fueling panic and fear among people [12]. This is an unproven but testable hypothesis, because users of social media use the platforms to express their emotions, feelings, and thoughts, which can be a valuable source of data for researching mental health [13].

ABC News reported a poll claiming that in the age of social media, anxiety about the coronavirus spreads faster than the virus itself, resulting in public panic worldwide [14]. On the other hand, social media is also a practical platform for the spreading of public health messages to audiences [15].

Brewer on BBC News [16] posits that hearing a lot of information and news about COVID-19 has affected the public and created panic, causing people to live with anxiety. Similarly, Rothschild and Fischer [17] claimed that social media is spreading fear and panic among social media users. Correspondingly, in the discussion on social media, Cellan-Jones [18] stated that people depend on social media to gain information and facts about COVID-19, as some countries use filters, which is why social media gives some information but not all the facts.

After COVID-19 appeared and was transmitted to other countries outside of Mainland China, people turned to social media to know more about the virus. According to Molla [19], in just 24 hours, there were 19 million mentions of COVID-19 across social media and news sites worldwide.

The mass media has been called on to take responsibility for providing correct information and aiding comprehension among citizens [20]. Frenkel et al [21] reports that after the WHO claimed that social media companies were fueling misinformation on COVID-19 worldwide, some social media companies tried to remove false information from their platforms.

Victor [22] claims that in today's digital age, Chinese citizens could not get enough facts about COVID-19, which is why they depended on social media and widely shared their information, photos, and videos, sometimes inaccurately. Likewise, in India, the government has asked top social media companies like Facebook, YouTube, TikTok, ShareChat, and Twitter to stop publishing misinformation, as it creates panic among people. Similarly, Emmott [23] noted that, according to a European Union document, Russian media has published a "significant disinformation campaign" about the COVID-19 outbreak to create panic among the public in Western countries.

In a contemporary discussion on the effects of media, one researcher [24] stated that in some countries, social media impacted the buying crisis, when many people tried to buy toilet paper and other items because of the spreading fear of COVID-19 on social media. According to the newspaper The Star [25], social media is responsible for much of the panic surrounding COVID-19, internationally leading to a situation where social media companies tried to eradicate posts about COVID-19 from their platforms. Furthermore, Devlin [26] stated that people saw posts of empty shops on social media, which created panic related to food shortages. Additionally, Kent [27] noted that social media gave everyone the chance to share information with everyone else, which is why people posted on social media as soon as they heard something about COVID-19. In addition, it is noted that publishing inaccurate information on social media networks about the spread of

diseases will have a negative impact on public health and people's mental health [28]. The public sphere in the 21st century has undergone a transformation generated by the adoption of online communication technologies. New media has become an important source of health information and a platform for discussing personal experiences, opinions, and concerns regarding health, illnesses, and treatment [28]. Similarly, Dillon [29] noted that people spend a lot of time on social media and may see cases of panic buying in various countries during the COVID-19 pandemic, which can spread panic further. In addition, El-Terk [30] showed that nowadays everyone is an expert because everyone tries to have a voice and send a message about COVID-19. Correspondingly, Garrett [31] explained that we gave power to social media to create fear about COVID-19, as we all publish panic-inducing information and it circulates.

Merchant and Lurie [32] found that at present, due to the development of social media, many methods of communicating and disseminating information and news are available to the public. These are fast and effective and can spread true information as well as misinformation. In addition, La et al [33] said that many countries did not circulate information to the public about the COVID-19 outbreak or were unable to provide the public with the information they needed; thus, people relied on the information they could find on social media. The Vietnamese case is a successful example of dealing with social media in the right way. The country's Ministry of Health created accounts on social media networks, and through those accounts, they published information about COVID-19 to the public.

Mian and Khan [34] argue that there has been a worldwide increase in the spread of fake news and misinformation about COVID-19, with misinformation such as the lab theory on the origin of the virus allegedly "originating" on social media. Correspondingly, Petric and others [35,36] believe that "media coverage has highlighted COVID-19 as a unique threat, rather than one of many, which has added to panic, stress." Depoux and others [37-40] determined that social media has played three main roles in the COVID-19 outbreak in most countries. First, facts about the outbreak were published on social media. Second, misinformation, fake news, and inaccurate information about the outbreak was published on social media. Third, social media created fear and panic about the outbreak worldwide.

Little or no evidence is available on the perception and impact of social media during this pandemic, in particular within non-Western communities such as Iraqui Kurdistan.

Methods

In this study we used a quantitative survey methodology to obtain data from Kurdish social media users. The questionnaire was prepared in the Kurdish language, and 516 social media users were sampled to collect the data. A descriptive content analysis was used to analyze the data. SPSS Version 25 (IBM Corp) was used to categorize and test the results. The social media users participated in a random online questionnaire, which aimed to determine the impact of social media on the spreading of panic about the COVID-19 outbreak, as well as social media's impact on people's mental health and the health crisis facing countries worldwide.

Results

Table 1 indicates that, of 516 study participants, 294 (57%) were male and 222 (43%) were female. In addition, most of the participants (n=336, 65.1%) were aged 18-35 years. Those who were 51 years and older made up only 6% (n=31) of the participants. The participants were divided into nine categories based on their scientific qualification. The most common scientific qualification was a Bachelor degree (n=261, 50.6%), while the least common one was a higher diploma (n=3, 0.6%).

The first question in this study asked participants "Which social media platform do you use to get news and information about COVID-19?" As shown in Table 2, the majority of participants (426/516, 82.6%) used Facebook to acquire information about COVID-19. The platforms TikTok, Skype, WeChat, and Myspace were among the lowest used for news and information. Facebook is at the top because it is the most popular social media platform used in the Kurdistan Region of Iraq.

The second question was "What news topics have you mostly heard/seen/read on social media during these three months of 2020?" As shown in Table 3, the highest proportion of participants (n=394, 76.4%) had heard, seen, or read health news (COVID-19), while the lowest proportion of participants had heard, seen, or read technology news (n=3, 0.6%). The survey results of the 516 participants show that the COVID-19 health crisis is affecting the type of news topics most commonly followed on social media.

Table 1. Sociodemographic variables of study participants (N=516).

Variables	Participants, n (%)
Gender	
Male	294 (56.9)
Female	222 (43.0)
Age (years)	
18-35	336 (65.1)
36-50	149 (28.9)
≥51	31 (6.0)
Scientific qualifications	
PhD	43 (8.3)
Master of Arts	85 (16.5)
Higher diploma	3 (0.6)
Bachelor	261 (50.6)
Diploma	65 (12.6)
High school	35 (6.8)
Secondary school	11 (2.1)
Primary school	7 (1.4)
Just reading and writing	6 (1.2)

Table 2. The social media platforms used to get news about the coronavirus disease.

Social media platforms	Participants (N=516), n (%)
Facebook	426 (82.6)
Instagram	33 (6.4)
Twitter	17 (3.3)
Snapchat	2 (0.4)
YouTube	10 (1.9)
TikTok	1 (0.2)
LinkedIn	6 (1.2)
WhatsApp	3 (0.6)
Telegram	4 (0.8)
Skype	1 (0.2)
Viber	9 (1.7)
LINE	2 (0.4)
WeChat	1 (0.2)
VKontakte (VK)	0 (0.0)
Badoo	0 (0.0)
Myspace	1 (0.2)

Table 3. The news topics classifications.

News topics	Participants (N=516), n (%)
Social news	14 (2.7)
Health news (COVID-19[a])	394 (76.4)
Technology news	3 (0.6)
Economic news	10 (1.9)
Sports news	4 (0.8)
Miscellaneous news	65 (12.6)
Political news	20 (3.9)
Cultural news	6 (1.2)

[a]COVID-19: coronavirus disease.

Cronbach alpha was used to determine the reliability of the study; its value was .825 and the validity was 0.753, indicating that the study questionnaire is highly reliable. Reliability refers to the accuracy, dependability, stability, and consistency of the research instrument. The recommended appropriate sample size is "approximately 200 individuals (or more) for a research" [41] indicating that the sample size of 516 respondents in this study was appropriate.

Table 4 shows the data on repeat distributions (mean, standard deviation, coefficient of variation, and relative importance) and indicates the explanatory variables that focus on Questions 3, 5, 6, and 8.

Table 5 shows the responses of 516 people to the question "If your answer to the sixth question is Yes, how did that fear affect you?" As shown in the table, 38.6% (n=199) of the participants were psychologically affected, while 36.0% (n=186) stated that they were not afraid. A minority stated that they were physically affected (n=9, 1.7%). The responses of the 516 participants showed that fear was primarily a psychological response that could cause a reduction in physical immunity, which is one of the reasons for poor outcomes when infected with COVID-19.

Table 4. Descriptive statistics of questions.

Questions	Value, mean (SD)	Coefficient of variation	Relative importance
Question 3: Do you think that publishing more news related to COVID-19[a] on social media has spread fear and panic among the people?	2.68 (0.63)	23.51	89.333
Question 5: Do you think the level of Kurdish pages, groups, and accounts on social media covering COVID-19 is good?	1.96 (0.88)	44.9	65.333
Question 6: Have you published any information and news related to COVID-19 on social media?	2.18 (0.93)	42.66	72.667
Question 8: Filters need to be set up for social media and a specific policy followed during humanitarian crises such as the spread of the COVID-19.	2.74 (0.62)	22.63	91.333
Total	2.39 (0.765)	33.425	79.667

[a]COVID-19: coronavirus disease.

Table 5. Impacts of fear on study participants (N=516).

Impact scale	Participants, n (%)
Psychological	199 (38.6)
Physical	9 (1.7)
Physical psyche	47 (9.1)
All of them	75 (14.6)
I was not afraid	186 (36.0)

Participants in this study were also asked, "Which category of information has had the most impact on creating panic on social media?" As shown in Table 6, many participants (n=137, 26.6%) answered "fake news about COVID-19," and 90 (17.4%) said it was "dissemination of the number of infections." In addition, 39 (7.6%) participants chose "dissemination of the number of deaths." This indicates that fake news and misinformation have an immediate and massive impact on individuals during this crisis, but also factual information such as the number of cases.

The responses to Questions 3, 6, and 8 (Table 7) indicate that more males than females responded yes, neutral, and no, but in Question 5, the rate of females was higher than males for the response neutral.

Table 6. Categories of information shared on social media.

Information	Participants (N=516), n (%)
Dissemination of the number of infections (A)	90 (17.4)
Dissemination of the death toll (B)	39 (7.6)
Dissemination of fear-inducing information about COVID-19[a] (C)	56 (10.9)
Publication of photos and videos of the cities and countries with a high number of cases (D)	78 (15.1)
Fake news about COVID-19 (E)	137 (26.6)
Dissemination of the number of infections (A) and dissemination of the death toll (B)	13 (2.5)
Dissemination of the number of infections (A) and dissemination of fear-inducing information about COVID-19 (C)	4 (0.8)
Dissemination of the number of infections (A) and publication of photos and videos of the cities and countries with a high number of cases (D)	9 (1.7)
Dissemination of the number of infections (A) and fake news about COVID-19 (E)	7 (1.4)
Dissemination of the death toll (B) and dissemination of fear-inducing information about COVID-19 (C)	3 (0.6)
Other	80 (15.9)

[a]COVID-19: coronavirus disease.

Table 7. Some questions according to the gender of participants (N=516).

Variables	Male, n (%)	Female, n (%)	Total, n (%)
Question 3: Do you think that publishing more news related to COVID-19[a] on social media has spread fear and panic among the people?			
No	25 (53.2)	22 (46.8)	47 (100.0)
Neutral	36 (51.4)	34 (48.6)	70 (100.0)
Yes	233 (58.4)	166 (41.6)	399 (100.0)
Question 5: Do you think the level of Kurdish pages, groups, and accounts on social media covering COVID-19 is good?			
No	144 (68.3)	67 (31.8)	211 (100.0)
Neutral	49 (43.4)	64 (56.3)	113 (100.0)
Yes	101 (52.6)	91 (47.4)	192 (100.0)
Question 6: Have you published any information and news related to COVID-19 on social media?			
No	133 (71.5)	73 (28.5)	186 (100.0)
Neutral	30 (60.0)	20 (40.0)	50 (100.0)
Yes	151 (53.9)	129 (46.1)	280 (100.0)
Question 8: Filters need to be set up for social media and a specific policy followed during humanitarian crises such as the spread of COVID-19.			
No	37 (75.5)	12 (24.5)	49 (100.0)
Neutral	22 (64.7)	12 (35.3)	34 (100.0)
Yes	235 (52.3)	198 (45.7)	433 (100.0)

[a]COVID-19: coronavirus disease.

According to the results shown in Table 8, the majority of Facebook users in this study were male (n=251, 58.9%) and 41.1% (n=175) were female. The majority of participants that were Instagram users were female (n=26, 78.8%), while a minority were male (n=7, 21.2%). Furthermore, of study participants that read economic news, 60% (n=6) were male and 40% (n=4) were female. Finally, of the study participants that read sports news, 75% (n=3) were male and 25% (n=1) female.

As shown in Table 9, most participants that reported using Facebook were 18-35 years of age (n=283, 66.4%), 124 (29.1%) were 36-50 years of age, and 19 (4.5%) were 51 years or older. The majority of participants that used Instagram were 18-35 years of age (n=28, 84.9%), and 5 (15.2%) of those that reported using Instagram for news were 36-50 years of age. Additionally,

of those that read economic news, 4 (40%) were 18-35 years of age, and 6 (60%) were 36-50 years of age.

As shown in Table 10, 111 (37.8%) males and 88 (39.6%) females felt psychological fear. Of all participants that said they felt psychological fear, 135 (40.2%) were 18-35 years of age, and 57 (38.3%) were 36-50 years of age.

Table 8. Accounting some questions according to gender of participants (N=516).

Variable	Gender		Total
	Male, n (%)	Female, n (%)	
Question 1: Which social media platform do you use to get news and information about COVID-19[a]?			
Facebook	251 (58.9)	175 (41.1)	426 (100.0)
Instagram	7 (21.2)	26 (78.8)	33 (100.0)
Twitter	10 (58.8)	7 (41.2)	17 (100.0)
Snapchat	0 (0.0)	2 (100.0)	2 (100.0)
YouTube	6 (60.0)	4 (40.0)	10 (100.0)
TikTok	0 (0.0)	1 (100.0)	1 (100.0)
LinkedIn	3 (50.0)	3 (50.0)	6 (100.0)
WhatsApp	3 (100.0)	0 (0.0)	3 (100.0)
Telegram	3 (75.0)	1 (25.0)	4 (100.0)
Skype	1 (100.0)	0 (0.0)	1 (100.0)
Viber	7 (77.8)	2 (22.2)	9 (100.0)
LINE	1 (50.0)	1 (50.0)	2 (100.0)
WeChat	1 (100.0)	0 (0.0)	1 (100.0)
Myspace	1 (100.0)	0 (0.0)	1 (100.0)
Question 2: What news topic have you primarily heard/seen/read on social media during these three months of 2020?			
Social news	12 (85.7)	2 (14.3)	14 (100.0)
Health news (COVID-19)	216 (54.8)	178 (45.2)	394 (100.0)
Technology news	2 (66.7)	1 (33.3)	3 (100.0)
Economic news	6 (60.0)	4 (40.0)	10 (100.0)
Sport news	3 (75.0)	1 (25.0)	4 (100.0)
Miscellaneous news	34 (52.3)	31 (47.7)	65 (100.0)
Political news	17 (85.0)	3 (15.0)	20 (100.0)
Cultural news	4 (66.7)	2 (33.3)	6 (100.0)

[a]COVID-19: coronavirus disease.

Table 9. Accounting some questions according to age of participants (N=516).

Variables	Age, n (%)			Total, n (%)
	18-35 years	36-50 years	≥51 years	
Question 1: Which social media platform do you use to get news and information about COVID-19[a]?				
Facebook	283 (66.4)	124 (29.1)	19 (4.5)	426 (100.0)
Instagram	28 (84.9)	5 (15.2)	0 (0.0)	33 (100.0)
Twitter	10 (58.8)	7 (41.2)	0 (0.0)	17 (100.0)
Snapchat	2 (100.0)	0 (0.0)	0 (0.0)	2 (100.0)
YouTube	4 (40.0)	4 (40.0)	2 (20.0)	10 (100.0)
TikTok	1 (100.0)	0 (0.0)	0 (0.0)	1 (100.0)
LinkedIn	3 (50.0)	2 (33.3)	1 (16.7)	6 (100.0)
WhatsApp	1 (33.3)	0 (0.0)	2 (66.7)	3 (100.0)
Telegram	1 (25.0)	1 (25.0)	2 (50.0)	4 (100.0)
Skype	0 (0.0)	1 (100.0)	0 (0.0)	1 (100.0)
Viber	2 (22.2)	3 (33.3)	4 (44.4)	9 (100.0)
LINE	0 (0.0)	1 (50.0)	1 (50.0)	2 (100.0)
WeChat	0 (0.0)	1 (100.0)	0 (0.0)	1 (100.0)
Myspace	1 (100.0)	0 (0.0)	0 (0.0)	1 (100.0)
Question 2: What news topics have you mostly heard/seen/read on social media during these three months of 2020?				
Social news	9 (64.3)	3 (21.4)	2 (14.3)	14 (100.0)
Health news (COVID-19)	266 (67.5)	112 (28.4)	16 (4.1)	394 (100.0)
Technology news	3 (100.0)	0 (0.0)	0 (0.0)	3 (100.0)
Economic news	4 (40.0)	6 (60.0)	0 (0.0)	10 (100.0)
Sport news	2 (50.0)	2 (50.0)	0 (0.0)	4 (100.0)
Miscellaneous news	41 (63.1)	17 (26.2)	7 (10.8)	65 (100.0)
Political news	8 (40.0)	6 (30.0)	6 (30.0)	20 (100.0)
Cultural news	3 (50.0)	3 (50.0)	0 (0.0)	6 (100.0)

[a]COVID-19: coronavirus disease.

Table 10. Variable description by age and gender.

Demographics	Variable					Total, n (%)
	Psychological, n (%)	Physical, n (%)	Psychological and physical, n (%)	All of them, n (%)	I was not afraid, n (%)	
Gender						
Male	111 (37.8)	5 (1.7)	24 (8.2)	42 (14.3)	112 (38.1)	294 (100.0)
Female	88 (39.6)	4 (1.8)	23 (10.4)	33 (14.9)	74 (33.3)	222 (100.0)
Combined	199 (38.7)	9 (1.7)	47 (9.1)	75 (14.6)	186 (36)	516 (100.0)
Age (years)						
18-35	135 (40.2)	6 (1.8)	36 (10.7)	43 (12.8)	116 (34.5)	336 (100.0)
36-50	57 (38.3)	2 (1.3)	9 (6.0)	23 (15.4)	58 (38.9)	149 (100.0)
≥51	7 (22.6)	1 (3.2)	2 (6.5)	9 (29.03)	12 (38.7)	31 (100.0)
Combined	199 (38.7)	9 (1.7)	47 (9.1)	75 (14.6)	186 (36.0)	516 (100.0)

It is noted from Table 11 that there is a significant positive statistical correlation between social media and the spreading of panic about COVID-19. The total variation is equal to 75.7%, which indicates that 75.7% of the variance of spreading panic

about COVID-19 has been explored in social media, and the other variables (24.3%) are due to random error. In other words, this illustrates that only 75.7% of the factors that affect spreading panic about COVID-19 are related to social media.

Table 11. Simple regression model analysis of a dependent variable (spreading panic about coronavirus disease) on the effects of social media on spreading panic about coronavirus disease and social media's impact on mental health in the Kurdistan Region of Iraq.

Model	Unstandardized coefficients		*t* test	*P* value	*R*	R^2	*F* test	*P* value
	B	SE						
Constant	0.4456	0.219	4.865	.001	.8701	.757	95.652	<.001
Social media	0.6458	0.0588	11.532	<.001	N/A[a]	N/A	N/A	N/A

[a]Not applicable.

Discussion

Overview

As media professionals working at a public university in the Kurdistan Region of Iraq, we conclude from the study results that social media has played a significant role in affecting the public during the COVID-19 crisis. The regression analysis of the study indicates that there is a significant positive statistical correlation (*R*=.8701) between social media and spreading panic about COVID-19. Moreover, we can see that it has had a psychological effect, primarily on the younger generation, where 40.2% (n=135/336) of the respondents aged 18-35 years were affected. People are gathering information from governmental sources that have eroded, and people are far more likely to get their information from social media than from any other sources. People are also unable to discern which information on social media is true and which is false, thus causing more panic and rumors about the true nature of the epidemic.

One could argue that the panic caused by widespread information about COVID-19 in the Kurdistan Region of Iraq is worse than the number of COVID-19 cases and will have a longer-lasting effect. It is important to communicate this to health professionals in the region and for media experts to work with these professionals to ensure that only well-vetted information is disseminated to the public. It is also important to engage the Ministry of Health and the Ministry of Education in this effort to be prepared for future epidemics or health situations. This pandemic has certainly helped the authors identify the need for educating consumers on health topics found through social media.

Limitations

There were various research limitations, most importantly these are self-reported data from self-selected participants, and the lockdown period was a constraint to gather more representative data. It was difficult to find participants who wished to participate in this study.

Conclusions

As media experts and educators, we have an important role to play both now and in the future of Kurdistan. We must work to educate media consumers on what constitutes good and reliable information and how to critically think through this information. Since younger people are also consuming information from social media and then spreading it to their family and friends, universities are ideal places to design courses and symposiums that can help students and faculty discern how to search for, find, and evaluate health information in the case of an epidemic or pandemic.

Acknowledgments

With the outbreak of this pandemic, the world has suffered from COVID-19. Our interest lies in the heart of developing knowledge; thus, the idea for this study was born. Here, the authors would like to extend our gratitude to everyone who was part of our research community, as their views have highly enriched our study.

In addition, we extend our acknowledgment to Dr Paiman Ahmad for reviewing this manuscript prior to submission and at the final editing phase. No funding was provided for conducting this study.

Conflicts of Interest

None declared.

References

1. World Health Organization. Coronavirus. 2020. URL: https://www.who.int/health-topics/coronavirus#tab=tab_2 [accessed 2020-04-07]
2. NewScientist. Covid-19: the disease caused by a kind of coronavirus which first originated in Wuhan, China in late 2019 URL: https://www.newscientist.com/term/covid-19/ [accessed 2020-05-06]
3. Hernández-García I, Giménez-Júlvez T. Assessment of Health Information About COVID-19 Prevention on the Internet: Infodemiological Study. JMIR Public Health Surveill 2020 Apr 01;6(2):e18717 [FREE Full text] [doi: 10.2196/18717] [Medline: 32217507]

4. Pappot N, Taarnhøj GA, Pappot H. Telemedicine and e-Health Solutions for COVID-19: Patients' Perspective. Telemed J E Health 2020 Apr 24:1-3. [doi: 10.1089/tmj.2020.0099] [Medline: 32329654]
5. McGowan BS, Wasko M, Vartabedian BS, Miller RS, Freiherr DD, Abdolrasulnia M. Understanding the factors that influence the adoption and meaningful use of social media by physicians to share medical information. J Med Internet Res 2012 Sep 24;14(5):e117 [FREE Full text] [doi: 10.2196/jmir.2138] [Medline: 23006336]
6. Woo H, Cho Y, Shim E, Lee J, Lee C, Kim SH. Estimating Influenza Outbreaks Using Both Search Engine Query Data and Social Media Data in South Korea. J Med Internet Res 2016 Jul 04;18(7):e177 [FREE Full text] [doi: 10.2196/jmir.4955] [Medline: 27377323]
7. Kemp S. Digital 2020: Iraq. 2020. URL: https://datareportal.com/reports/digital-2020-iraq [accessed 2020-02-18]
8. Mavragani A. Infodemiology and Infoveillance: Scoping Review. J Med Internet Res 2020 Apr 28;22(4):e16206 [FREE Full text] [doi: 10.2196/16206] [Medline: 32310818]
9. Kurdistan Regional Government. Coronavirus (COVID-19) information and advice. 2020. URL: https://gov.krd/coronavirus-en/ [accessed 2020-04-10]
10. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
11. JMIR E-collection "Infoveillance, Infodemiology and Digital Disease Surveillance". URL: https://publichealth.jmir.org/themes/304 [accessed 2020-03-14]
12. Hao K, Basu T. The coronavirus is the first true social-media infodemic. 2020. URL: https://www.technologyreview.com/s/615184/the-coronavirus-is-the-first-true-social-media-infodemic/ [accessed 2020-02-12]
13. Wongkoblap A, Vadillo MA, Curcin V. Researching Mental Health Disorders in the Era of Social Media: Systematic Review. J Med Internet Res 2017 Jun 29;19(6):e228 [FREE Full text] [doi: 10.2196/jmir.7215] [Medline: 28663166]
14. Muwahed J. Coronavirus pandemic goes viral in the age of social media, sparking anxiety. URL: https://tinyurl.com/ybnms2se [accessed 2020-03-14]
15. Gough A, Hunter RF, Ajao O, Jurek A, McKeown G, Hong J, et al. Tweet for Behavior Change: Using Social Media for the Dissemination of Public Health Messages. JMIR Public Health Surveill 2017 Mar 23;3(1):e14 [FREE Full text] [doi: 10.2196/publichealth.6313] [Medline: 28336503]
16. Brewer K. Coronavirus: How to protect your mental health. 2020. URL: https://www.bbc.com/news/health-51873799 [accessed 2020-03-16]
17. Rothschild N. Coronavirus panic sells as alarmist information flies on social media. 2020. URL: https://www.axios.com/coronavirus-social-media-b56326b6-ab16-4c8a-bc86-e29b06e5ab2b.html [accessed 2020-03-03]
18. Cellan-Jones R. Tech Tent: Is social media spreading the virus?. 2020. URL: https://www.bbc.com/news/technology-51510196 [accessed 2020-02-14]
19. Molla R. How coronavirus took over social media. 2020. URL: https://tinyurl.com/ycwtmx3u [accessed 2020-03-12]
20. Gesser-Edelsburg A, Diamant A, Hijazi R, Mesch GS. Correcting misinformation by health organizations during measles outbreaks: A controlled experiment. PLoS One 2018 Dec 19;13(12):e0209505 [FREE Full text] [doi: 10.1371/journal.pone.0209505] [Medline: 30566485]
21. Frenkel S, Alba D, Zhong R. Surge of Virus Misinformation Stumps Facebook Twitter. 2020. URL: https://www.nytimes.com/2020/03/08/technology/coronavirus-misinformation-social-media.html [accessed 2020-03-08]
22. Victor D. Panic and Criticism Spread on Chinese Social Media Over Coronavirus. 2020. URL: https://www.nytimes.com/2020/01/24/world/asia/china-social-media-coronavirus.html [accessed 2020-01-24]
23. Emmott R. Russia deploying coronavirus disinformation to sow panic in West, EU document says. 2020. URL: https://tinyurl.com/yx42jzyh [accessed 2020-03-18]
24. Hernandez S. Panic spreads fear especially on social media, local doctor says. 2020. URL: https://tinyurl.com/y736nqo7 [accessed 2020-03-20]
25. The Star. Coronavirus and how social media firms are combating the spread of misinformation about the disease. 2020. URL: https://tinyurl.com/y85jagfn [accessed 2020-03-05]
26. Devlin H. Don't let coronavirus tip society into panic, say psychologists. 2020. URL: https://www.theguardian.com/science/2020/mar/06/dont-let-coronavirus-tip-society-into-panic-say-psychologists [accessed 2020-03-06]
27. Kent S. Fear spreads like a virus: How coronavirus panic is taking a toll on NJ's mental health. 2020. URL: https://tinyurl.com/ybwxk6o3 [accessed 2020-03-08]
28. Gold H. Inside the WHO's fight to stop false information about coronavirus from spreading. 2020. URL: https://www.wktv.com/content/news/568519502.html [accessed 2020-03-05]
29. Dillon C. Coronavirus: The psychology behind panic-buying. 2020. URL: https://tinyurl.com/y9orfutq [accessed 2020-03-11]
30. El-Terk N. Toilet paper, canned food: What explains coronavirus panic buying. 2020. URL: https://tinyurl.com/ww23bb3 [accessed 2020-03-13]
31. Garrett O. A Case of Modern Mass Hysteria? The Coronavirus. 2020. URL: https://tinyurl.com/yarx6j7r [accessed 2020-02-13]
32. Merchant RM, Lurie N. Social Media and Emergency Preparedness in Response to Novel Coronavirus. JAMA 2020 Mar 23:29-31. [doi: 10.1001/jama.2020.4469] [Medline: 32202611]

33. La V, Pham T, Ho M, Nguyen M, P. Nguyen K, Vuong T, et al. Policy Response, Social Media and Science Journalism for the Sustainability of the Public Health System Amid the COVID-19 Outbreak: The Vietnam Lessons. Sustainability 2020 Apr 07;12(7):2931. [doi: 10.3390/su12072931]
34. Mian A, Khan S. Coronavirus: the spread of misinformation. BMC Med 2020 Mar 18;18(1):89-82 [FREE Full text] [doi: 10.1186/s12916-020-01556-3] [Medline: 32188445]
35. Petric D. Negative mental health effects of COVID-19 pandemic panic. 2020. URL: https://tinyurl.com/y8mqkpde [accessed 2020-04-06]
36. Shimizu K. 2019-nCoV, fake news, and racism. The Lancet 2020 Feb;395(10225):685-686. [doi: 10.1016/s0140-6736(20)30357-3]
37. Depoux A, Martin S, Karafillakis E, Bsd RP, Wilder-Smith A, Larson H. The pandemic of social media panic travels faster than the COVID-19 outbreak. J Travel Med 2020 Mar 03:1-4. [doi: 10.1093/jtm/taaa031] [Medline: 32125413]
38. Saud M, Mashud M, Ida R. Persistence of Social Media on Political Activism and Engagement among Indonesian & Pakistani youths. IJWBC 2020;16(1):1 [FREE Full text] [doi: 10.1504/ijwbc.2020.10028448]
39. Mahmood QK, Saud M. Civic Participation and Facebook: A Cross Sectional Survey of Pakistani University Students. 2018 Presented at: 4th International Conference on Contemporary Social and Political Affairs (ICoCSPA) 2018; 13 August; Surabaya p. 59-65.
40. Saud M. Social Networking Sites and Deviance among Youth in Islamabad, Pakistan. EJBS 2015 May 11:48-58. [doi: 10.33422/ejbs.2018.05.85]
41. Dawson C. Activity 22: Knowing about Probability Samples. In: 100 Activities for Teaching Research Methods. London: SAGE Publications Ltd; 2016.

Abbreviations

COVID-19: coronavirus disease
KRG: Kurdistan Regional Government
WHO: World Health Organization

Edited by G Eysenbach; submitted 22.04.20; peer-reviewed by N Hamassed, H Kakashekh, M Saud, MA Bahrami; comments to author 30.04.20; revised version received 04.05.20; accepted 04.05.20; published 19.05.20.

Please cite as:
Ahmad AR, Murad HR
The Impact of Social Media on Panic During the COVID-19 Pandemic in Iraqi Kurdistan: Online Questionnaire Study
J Med Internet Res 2020;22(5):e19556
URL: http://www.jmir.org/2020/5/e19556/
doi:10.2196/19556
PMID:32369026

Original Paper

Measuring the Outreach Efforts of Public Health Authorities and the Public Response on Facebook During the COVID-19 Pandemic in Early 2020: Cross-Country Comparison

Aravind Sesagiri Raamkumar[1], BTech, MSc, PhD; Soon Guan Tan[1], BSc; Hwee Lin Wee[1,2], BSc, PhD

[1]Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore

[2]Department of Pharmacy, National University of Singapore, Singapore, Singapore

Corresponding Author:
Hwee Lin Wee, BSc, PhD
Saw Swee Hock School of Public Health
National University of Singapore
MD1 #10-01
12 Science Drive 2, National University of Singapore
Singapore, 117549
Singapore
Phone: 65 65164975
Email: ephwhl@nus.edu.sg

Abstract

Background: The coronavirus disease (COVID-19) pandemic presents one of the most challenging global crises at the dawn of a new decade. Public health authorities (PHAs) are increasingly adopting the use of social media such as Facebook to rapidly communicate and disseminate pandemic response measures to the public. Understanding of communication strategies across different PHAs and examining the public response on the social media landscapes can help improve practices for disseminating information to the public.

Objective: This study aims to examine COVID-19-related outreach efforts of PHAs in Singapore, the United States, and England, and the corresponding public response to these outreach efforts on Facebook.

Methods: Posts and comments from the Facebook pages of the Ministry of Health (MOH) in Singapore, the Centers for Disease Control and Prevention (CDC) in the United States, and Public Health England (PHE) in England were extracted from January 1, 2019, to March 18, 2020. Posts published before January 1, 2020, were categorized as pre-COVID-19, while the remaining posts were categorized as peri-COVID-19 posts. COVID-19-related posts were identified and classified into themes. Metrics used for measuring outreach and engagement were frequency, mean posts per day (PPD), mean reactions per post, mean shares per post, and mean comments per post. Responses to the COVID-19 posts were measured using frequency, mean sentiment polarity, positive to negative sentiments ratio (PNSR), and positive to negative emotions ratio (PNER). Toxicity in comments were identified and analyzed using frequency, mean likes per toxic comment, and mean replies per toxic comment. Trend analysis was performed to examine how the metrics varied with key events such as when COVID-19 was declared a pandemic.

Results: The MOH published more COVID-19 posts (n=271; mean PPD 5.0) compared to the CDC (n=94; mean PPD 2.2) and PHE (n=45; mean PPD 1.4). The mean number of comments per COVID-19 post was highest for the CDC (mean CPP 255.3) compared to the MOH (mean CPP 15.6) and PHE (mean CPP 12.5). Six major themes were identified, with posts about prevention and safety measures and situation updates being prevalent across the three PHAs. The themes of the MOH's posts were diverse, while the CDC and PHE posts focused on a few themes. Overall, response sentiments for the MOH posts (PNSR 0.94) were more favorable compared to response sentiments for the CDC (PNSR 0.57) and PHE (PNSR 0.55) posts. Toxic comments were rare (0.01%) across all PHAs.

Conclusions: PHAs' extent of Facebook use for outreach purposes during the COVID-19 pandemic varied among the three PHAs, highlighting the strategies and approaches that other PHAs can potentially adopt. Our study showed that social media analysis was capable of providing insights about the communication strategies of PHAs during disease outbreaks.

(J Med Internet Res 2020;22(5):e19334) doi:10.2196/19334

KEYWORDS

COVID-19; sentiment analysis; emotion analysis; public health authorities; infectious disease; outbreak; public engagement; social media; public health; virus

Introduction

Background

The coronavirus disease (COVID-19) was first identified in Wuhan, China in December 2019. It has since spread to 210 countries and territories, infecting 1,697,356 people and causing 102,667 deaths as of April 11, 2020 [1]. Compared to the previous outbreaks of severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome, COVID-19 has caused more infections and deaths, spreading from an infected person to 2-2.5 people on average [2]. Most countries started reporting infections by the second half of January 2020. The United States reported its first case on January 20, 2020 [3], while England, under the United Kingdom, reported its first cases on January 31, 2020 [4]. Singapore reported its first case on January 23, 2020 [5]. In an effort to contain the COVID-19 pandemic in Singapore, multiple interventions have been implemented on both societal and health care system levels [6], and the country shifted rapidly to Disease Outbreak Response System Condition (DORSCON) orange, the second-highest level of alert for disease outbreaks in Singapore, on February 7, 2020 [7], just 15 days after the first case of COVID-19 infection was confirmed. The World Health Organization (WHO) declared this disease as a pandemic on March 11, 2020, and is unable to ascertain the duration of the pandemic [8].

Countries such as Singapore, Taiwan, and South Korea have taken the necessary precautions to handle this pandemic within their borders in January when the pandemic was largely confined to China [9]. On the international scene, Singapore has received accolades from the WHO and several world leaders praising our efforts in containing the disease [10,11]. The Dean of the Saw Swee Hock School of Public Health, National University of Singapore, has also been invited by several overseas universities to share Singapore's experience in combating the COVID-19 pandemic [12,13]. Unfortunately, precautionary measures have been reported to be found wanting in other countries such as the United States and the United Kingdom [14,15]. It is to be noted that such delays in preparation for epidemics have also been seen in the past with Zika, influenza A virus subtype H1N1, and Ebola [16].

Social Media Use During Pandemics

Effective risk communication is essential in directing the public to adopt certain desired behaviors such as social distancing and good hygiene habits in times of pandemics. Transparent and consistent communication amidst the uncertainty of the pandemic is also crucial in maintaining public confidence and trust [17,18]. Traditionally, the government and public health authorities (PHAs) relied on websites, news media, print press, and television as main platforms for the dissemination of pandemic-related news and information to the public. In contrast to the 2003 SARS and 2009 H1N1 pandemic, present-day media landscapes worldwide have evolved significantly, with a greater presence of social media and alternative local and overseas media outlets [19]. The advent of social media platforms such as Facebook and Twitter facilitated the instantaneous sharing of information during pandemics for both the health authorities and the general public. With widespread social media use and the participatory web, PHAs must understand that health risk communication is no longer a linear process [20]. The public can voice their sentiments and comments on the actions undertaken by the government as events related to the pandemic unfold. The public themselves are also involved in content creation through blogs and forums. Citizen journalism and propagation of information pertaining to a pandemic is made possible within their social networks.

Existing research on social media has explored epidemics and pandemics such as Zika [21-23], H1N1 [24], and Ebola [25]. The scope of these studies includes descriptive analysis of posting frequency [21], thematic analysis of post content [26], sentiment analysis of posts [23], and social network analysis [24]. Although the WHO has put forth guidelines for emergency risk communication during epidemics [27], countries may adopt different strategies when conveying health risks across social media platforms. Currently, there is a lack of studies that compare the social media outreach efforts of PHAs from different countries and corresponding responses and interactions by the general public. Such studies might offer rich insights on how effectively platforms such as Facebook could be used for risk communication.

This Study

Amid the uncertainty of a health threat such as COVID-19, the public have a greater demand for real time, transparent, and consistent messaging. Government agencies run the risk of losing the centralized control of the risk communication process if they do not act swiftly to public sentiment and dispel falsehoods and misinformation [18,28]. A confluence of factors could lead to unintended behavioral outcomes among the public in the ongoing COVID-19 pandemic. The mismatch in perceived threats as well as costs and benefits of certain health behaviors communicated by either mainstream media, government authorities, or alternative media could result in a distorted understanding among the general public. Hence, it is crucial to understand how the prevailing sentiments and narratives about the pandemic were conveyed through the different communication channels and how it was received by the general public who have access to these channels. This will highlight the trigger points, allowing health authorities to fine-tune messaging along the course of the pandemic to allay public fear and panic.

Hence, in this study, we seek to answer four research questions related to Facebook use during a pandemic. First, how frequently do the PHAs of Singapore, the United States, and England use Facebook for risk communication? Second, what were the primary themes of the COVID-19-related posts by PHAs? Third, what are the Facebook followers' sentiments and emotions in response to these COVID-19-related posts by PHAs? Fourth,

how common are toxic comments that may incite public unrest, and do these toxic comments gain traction? We have selected Singapore, the United States, and England for this study, as we intend to look at the findings from a cross-country perspective and these are developed countries that have English as their official language.

Methods

Data Extraction

Data for this study were extracted from three Facebook pages using the tool Facepager [29] on March 19, 2020. The three Facebook pages are officially managed by the Ministry of Health (MOH), Singapore [30], the Centers for Disease Control and Prevention (CDC) in the United States [31], and Public Health England (PHE) in England [32]. As of April 4, 2020, the followers count of the MOH, the CDC, and PHE are 212,453, 2,636,072, and 336,935, respectively. Extracted data include posts by PHAs, comments from Facebook users, and their corresponding reactions, a feature in Facebook where users can interact with a Facebook status update, article, or a photo or video using one of six emotional reactions: Like, Love, Haha, Wow, Sad, and Angry. Contents posted between January 1, 2019, and March 18, 2020, were analyzed. Posts before January 1, 2020, were considered pre-COVID-19, and posts after January 1, 2020, were considered peri-COVID-19. We selected January 1, 2020, as the starting point for peri-COVID-19 since the statement on "Precautionary Measures in Response to Severe Pneumonia Cases in Wuhan, China," was issued by the MOH on January 2, 2020 [33]. COVID-19-related posts were filtered out by manually scanning through the textual content of the posts.

Data Analysis

Extent of Facebook Use

To examine the extent of Facebook use by the MOH, the CDC, and PHE, we calculated the average number of daily posts and compared pre-COVID-19 and peri-COVID-19 phases. To identify how specific events may influence the extent of Facebook use, we related the number of daily posts to the key dates on which the three countries reported their first COVID-19 cases or declared the outbreak as a national-level pandemic through the average number of posts per day (PPD) measure. To evaluate the extent of public engagement with the Facebook posts, we calculated the average number of reactions per post (RPP), average number of shares per post (SPP), and average number of comments per post (CPP) for pre- and peri-COVID-19 periods but focusing on COVID-19-related posts for the peri-COVID-19 period.

Thematic Analysis of COVID-19 Posts

The prevalent theme of each COVID-19 post was identified using a process involving two coders. First, the principal coder reviewed the contents of the COVID-19 posts by PHAs and assigned the relevant themes [34]. Next, another coder reviewed and confirmed the themes assigned by the first coder. The content of the posts were first screened through and condensed into short units. The predominant theme conveyed in the post (ie, the theme taking the larger proportion of the message) was assigned to posts with more than one theme. The list of themes included *situation update, preventive measures, appreciation, public reassurance, disease information, falsehood correction, face mask, research, testing and diagnosis, and miscellaneous*. For each of these themes, the number of constituent posts and percentage of these posts to the total number of posts were reported.

Sentiment and Emotion Analyses of Comments to COVID-19 Posts

To gain insights into the reactions and comments by Facebook users on the posts by PHAs, we conducted sentiment and emotion analyses on comments written in the English language. The sentiment polarity (SP) score for each comment was identified using the Vader algorithm [35] since the algorithm has been specifically conceptualized for ascertaining the sentiment in short texts (eg, user comments on Facebook or tweets from Twitter). The SP score value ranges from –1 to 1 and was classified into five categories: *verynegative (–1<score<–0.5), negative (–0.5<score<0), neutral (score=0), positive (0<score<0.5), or very positive (0.5<score<1).*

For emotion analysis, we went beyond the emotions available as Facebook reactions. We adopted the eight emotions put forth in the theory of emotion [36] and classified the emotions of the comments as: *anger, fear, sadness, disgust, surprise, anticipation, trust,* or *joy*. Among the eight emotions, *trust* and *joy* are positive emotions, while *anger*, *sadness*, *fear,* and *disgust* are considered negative emotions. *Surprise* and *anticipation* can be either positive or negative depending on the context, hence not included in either of the two categories. The emotions conveyed in the comments were identified with the help of the DeepMoji algorithm [37] using a two-step process. First, the comments were analyzed using the DeepMoji algorithm, which recommended emojis based on the textual content. Second, the emotion that was mapped to the first-ranked emoji was considered as the emotion for the comment. The table in Multimedia Appendix 1 lists the mapping between the emojis and the corresponding emotions.

The total number of comments, CPP, SP scores, positive to negative sentiments ratio (PNSR) and positive to negative emotions ratio (PNER) were reported. PNSR and PNER are two valid measures that have been used in prior studies for sentiment and emotion analysis in texts [38-40]. The number of comments and the SP scores were plotted alongside the dates of key events.

Identification of Toxicity in Comments of COVID-19 Posts

Besides the general sentiment and emotion analyses, we identified toxic comments that may warrant intervention. Toxicity is defined as "the usage of rude, disrespectful, or unreasonable language that will likely provoke or make another user leave a discussion" [41]. In this study, the Perspective application program interface (API) service of Google [42] was used to measure the toxicity of comments. The toxicity score ranges between 0 (nontoxicity) and 1 (full toxicity). We further categorized the comments into a dichotomous variable, where comments with a toxicity score greater than or equal to 0.75 are

toxic and comments with toxicity scores less than 0.75 are nontoxic. We examined the number of likes and replies per toxic comment to determine if such toxic comments may have any ripple effect. We compared this to the number of likes and replies per nontoxic comment as a reference.

Results

COVID-19 Outreach Efforts of PHAs and Public Engagement

In Table 1 and 2, the aggregated statistics related to the Facebook posts from the MOH, the CDC, and PHE are listed along with the public engagement metrics. In Figure 1, the COVID-19 posts daily count values are plotted in a line graph. In addition, the key dates on which the three countries reported their first COVID-19 cases and declared the outbreak as a national-level pandemic (in national terminology) are highlighted in the figure.

In the pre-COVID-19 phase (January 1, 2019, to December 31, 2019), the CDC had an average of 2 daily posts, while the MOH and PHE had an average of 1 daily post (Table 1). The MOH stepped up on the use of Facebook for public engagement during COVID-19. As of March 18, 2020, the MOH had published 304 posts peri-COVID-19 (January 1, 2020, to March 18, 2020), which was nearly a threefold increase in their PPD from 1.35 to 4.34. The MOH had the highest frequency of postings among the three PHAs with 304 posts, and the majority were related to COVID-19. The mean number of daily posts was 4.3, ranging from 1 to 15 posts daily. February was the most active month with 132 COVID-19 posts. Prior to the first locally confirmed case in Singapore on January 23, 2020, there was a limited number of daily posts from the MOH (ranging from 1 to 3 daily posts). However, the number of daily posts from the MOH increased to 7 two days after the first confirmed case in Singapore, and there have been at least 2 PPD ever since.

Of the total CDC posts, less than half were related to COVID-19. It published the highest number of COVID-19 posts in March, with 53 posts in the first 18 days of the month. The highest number of posts published on a single day was on March 8, 2020, with 5 posts. The CDC started to post at least 1 COVID-19 post every day starting on February 25, 2020. The number of daily posts remained at low levels even after the United States had declared a level 1 emergency on January 30, 2020. The average number of daily posts increased only after the United States declared COVID-19 as a level 1 emergency on January 30, 2020, and rose again after the WHO declared a pandemic on March 11, 2020.

PHE had the lowest posting frequency with only about half of posts being related to COVID-19 posts. Similar to the MOH, February was the most active month with 19 COVID-19 posts for PHE, and the highest number of posts on a single day (n=3) was done on 3 days: January 24, March 4, and March 13, 2020. Despite the United Kingdom declaring COVID-19 as a level 4 incident on March 3, 2020, there were days when PHE had zero COVID-19 posts. PHE started posting at least one COVID-19 post every day on March 13, 2020.

Compared to the pre-COVID-19 phase (Table 2), there was also a considerable increase in the public engagement metrics in the peri-COVID-19 time period. For instance, the MOH had a 7-fold increase in CPP, with a higher average number of people commenting in 2020 compared to 2019. The CDC saw a 9-fold increase in the mean RPP from 2019 to 2020 and close to a 10-fold increase in mean SPP from 2019 to 2020. In the case of PHE, the biggest rise is seen for SPP, with nearly a 5-fold increase from 2019 to 2020.

We have also noted that Facebook users who read peri-COVID-19 posts from the MOH were more likely to react to the post than to share or comment on the post, while Facebook users who read posts from the CDC and PHE were more likely to share the posts than to react or comment on the post. This observation was consistent in both pre- and peri-COVID-19 periods.

Table 1. Summary of COVID-19 Facebook outreach by the MOH, the CDC, and PHE during the periods of pre-COVID-19 (January 1, 2019, to December 31, 2019) and peri-COVID-19 (January 1, 2020, to March 18, 2020).

Agency	Outreach effort (pre-COVID-19[a])		Outreach effort (peri-COVID-19)			
	Total posts, n	PPD[b], mean (SD)	Total posts, n	PPD, mean (SD)	COVID-19 posts, n (%)	COVID-19 PPD Since first reported case in each country, mean (SD)
MOH[c]	192	1.4 (1.09)	304	4.3 (3.5)	271 (89.1)	5.0 (3.6)
CDC[d]	599	2.1 (0.99)	232	2.1 (1.1)	94 (40.5)	2.2 (1.1)
PHE[e]	346	1.3 (0.56)	87	1.6 (0.8)	45 (51.7)	1.4 (0.7)

[a]COVID-19: coronavirus disease.

[b]PPD: posts per day.

[c]MOH: Ministry of Health.

[d]CDC: Centers for Disease Control and Prevention.

[e]PHE: Public Health England.

Table 2. Summary of COVID-19 Facebook engagement by the MOH, the CDC, and PHE during the periods of pre-COVID-19 (January 1, 2019, to December 31, 2019) and peri-COVID-19 (January 1, 2020, to March 18, 2020).

Agency	Public engagement (pre-COVID-19[a])			Public engagement (peri-COVID-19)		
	RPP[b], mean (SD)	SPP[c], mean (SD)	CPP[d], mean (SD)	RPP_C19[e], mean (SD)	SPP_C19[f], mean (SD)	CPP_C19[g], mean (SD)
MOH[h]	34.3 (26.3)	20.1 (100.3)	2.2 (4.5)	188.9 (201.4)	84.6 (279.7)	15.6 (20.7)
CDC[i]	230.7 (203.4)	240.9 (697.8)	43.1 (72.7)	2128.2 (4864.9)	2373.8 (3485.8)	255.3 (298.3)
PHE[j]	52.4 (68.7)	102.6 (212.5)	4.3 (6.7)	101.5 (89.8)	478.9 (568.1)	12.5 (11.5)

[a]COVID-19: coronavirus disease.

[b]RPP: reactions per post.

[c]SPP: shares per post.

[d]CPP: comments per post.

[e]RPP_C19: reactions to COVID-19 post.

[f]SPP_C19: shares per COVID-19 post.

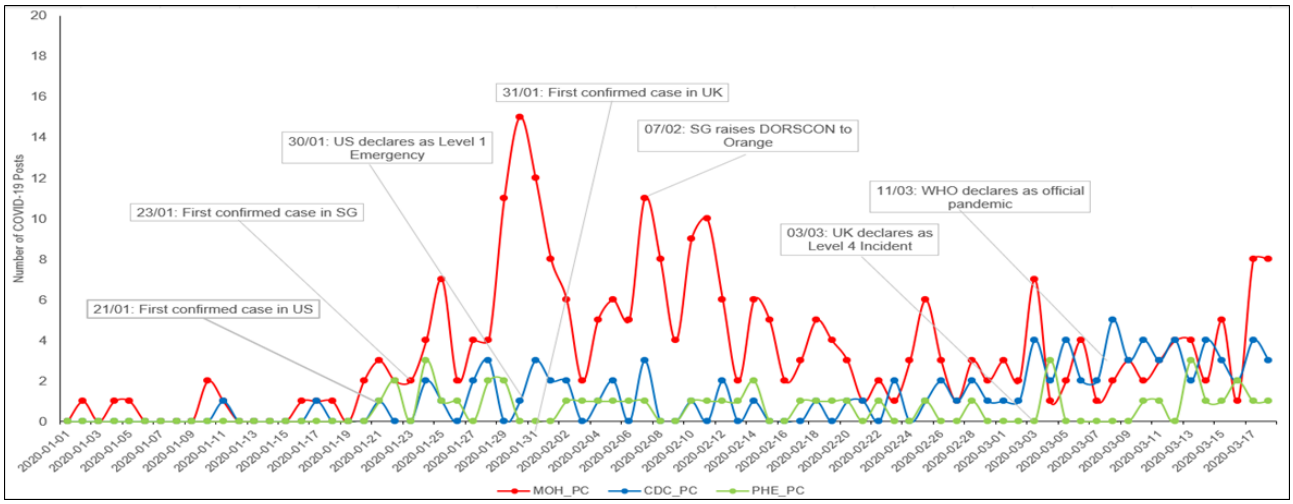
[g]CPP_C19: comments per COVID-19 post.

[h]MOH: Ministry of Health.

[i]CDC: Centers for Disease Control and Prevention.

[j]PHE: Public Health England.

Figure 1. COVID-19 posts frequency during the analysis period. CDC_PC: Centers for Disease Control and Prevention's number of posts; COVID-19: coronavirus disease; DORSCON: Disease Outbreak Response System Condition; MOH_PC: Ministry of Health's number of posts; PHE_PC: Public Health England number of posts; SG: Singapore; WHO: World Health Organization.



Thematic Analysis

The themes *appreciation*, *research*, *testing and diagnosis*, and *miscellaneous* were combined to the theme *others* to focus on six major themes. As shown in Table 3, the themes from the MOH are more diverse, with no theme exceeding 30% of the total posts. In contrast, the CDC and PHE posts were largely related to *preventive measures*. For the MOH, *situation update* and *preventive measures* were the top two themes. Interestingly, the CDC and PHE did not issue any post to correct false information, while the MOH issued 16 of such posts. There was also no post from PHE and only 2 posts from the CDC to reassure their Facebook followers, while the MOH issued 32 of such *public reassurance* posts.

Table 3. Thematic analysis of the public health authorities' coronavirus disease outreach efforts.

Theme	MOH[a] (n=271), n (%)	CDC[b] (n=94), n (%)	PHE[c] (n=45), n (%)
Preventive measures	60 (22.1)	50 (53.2)	18 (40.0)
Situation update	78 (28.8)	21 (22.3)	7 (15.6)
Disease information	16 (5.9)	17 (18.1)	17 (37.8)
Public reassurance	32 (11.8)	2 (2.1)	0 (0.0)
Falsehood correction	16 (5.9)	0 (0.0)	0 (0.0)
Others	69 (25.5)	4 (4.3)	3 (6.7)

[a]MOH: Ministry of Health.

[b]CDC: Centers for Disease Control and Prevention.

[c]PHE: Public Health England.

Sentiment and Emotion Analyses of Comments to COVID-19 Posts

The number of CPP was highest for the CDC, which was attributed to the high number of followers on their Facebook page (Table 4). Although the MOH had fewer number of followers than PHE, the number of comments received for the MOH's COVID-19 posts were 5 times more than PHE. This observation can also be attributed to the high number of COVID-19 posts for the MOH vs PHE's COVID-19 posts. Interestingly, the mean CPP of PHE was still higher than the MOH. The average SP scores of all three PHAs were close to the neutral sentiment mark of zero with only the MOH being slightly positive. Correspondingly, the PNSR and PNER of the MOH were much higher than the values for the CDC and PHE. However, since all these values were below 1, it is an indication that there were more negative sentiments and emotions conveyed in the comments. The CDC and PHE received predominantly negative comments from their followers based on the low PNSR and PNER values.

In Figure 2, the emotion categories are plotted against the sentiment categories, with a darker shade of the box reflecting a higher number of comments. We observed that most of the comments to the MOH posts were angry (n=1704, 33.9%), with 1215 being angry and negative and 489 being angry and very negative. Anger is also the most prevalent emotion for the CDC (n=12,634/42,470, 29.8%) and PHE (n=300/977, 30.7%) posts. For the MOH, the negative emotions (anger, disgust, fear, and sadness) account for 62.0% (n=3119/5032) of the comments, while positive emotions (trust and joy) accounted for 33.0% (n=1655) of the comments. In the case of the CDC and PHE, negative emotions accounted for 63.0% (n=26,716/42,470) and 66.2% (n=647/977) of the comments, respectively, and positive emotions accounted for 28.2% (n=11,987) and 26.0% (n=254) of comments, respectively. Since emotions have a direct effect on sentiments, the negative sentiments accounted for the majority of the comments (n=2431/5032, 48.3%; n=21,015/42,470, 49.5%; and n=491/977, 50.3% for the MOH, the CDC, and PHE, respectively). The MOH had a higher percentage of positive sentiments (n=1725/5032, 34.3%) compared to the CDC (n=12,256/42,470, 28.9%) and PHE (n=277/977, 28.4%).

The temporal trend analysis (Figure 3) provides more information compared to the snapshots provided in Table 4 and Figure 2. For instance, we observed that the number of comments increased significantly over time for the CDC posts, while the number of comments for the MOH posts appeared to have decreased over time. Among the three agencies, PHE had the highest degree of fluctuations in SP scores, with many negative and few positive spikes. The SP scores were mostly negative for the CDC posts, while SP scores were positive for the MOH on several occasions, which contributed to an average SP that tends toward neutral. The MOH had the highest number of days with positive sentiments, particularly the period between February 16, 2020, and March 5, 2020, which could be due to a relatively high number of appreciation posts (n=9) during that period.

Table 4. Emotion and sentiment analyses of COVID-19 Facebook comments.

Agency	COVID-19[a] posts, n	Comments, n	CPP[b], mean (SD)	SP[c], mean (SD)	PNSR[d], mean (SD)	PNER[e], mean (SD)
MOH[f]	271	5032	18.57 (30.04)	0.02 (0.25)	0.94 (2.11)	0.84 (1.79)
CDC[g]	94	42,470	451.81 (529.09)	–0.09 (0.06)	0.57 (0.16)	0.41 (0.12)
PHE[h]	45	977	21.71 (31.89)	–0.14 (0.26)	0.55 (0.49)	0.44 (0.54)

[a]COVID-19: coronavirus disease.

[b]CPP: comments per post.

[c]SP: sentiment polarity score.

[d]PNSR: positive to negative sentiments ratio.

[e]PNER: positive to negative emotions ratio.

[f]MOH: Ministry of Health.

[g]CDC: Centers for Disease Control and Prevention.

[h]PHE: Public Health England.

Figure 2. Sentiment and emotions heat map for coronavirus disease Facebook comments (left: Ministry of Health; middle: Centers for Disease Control and Prevention; right: Public Health England).

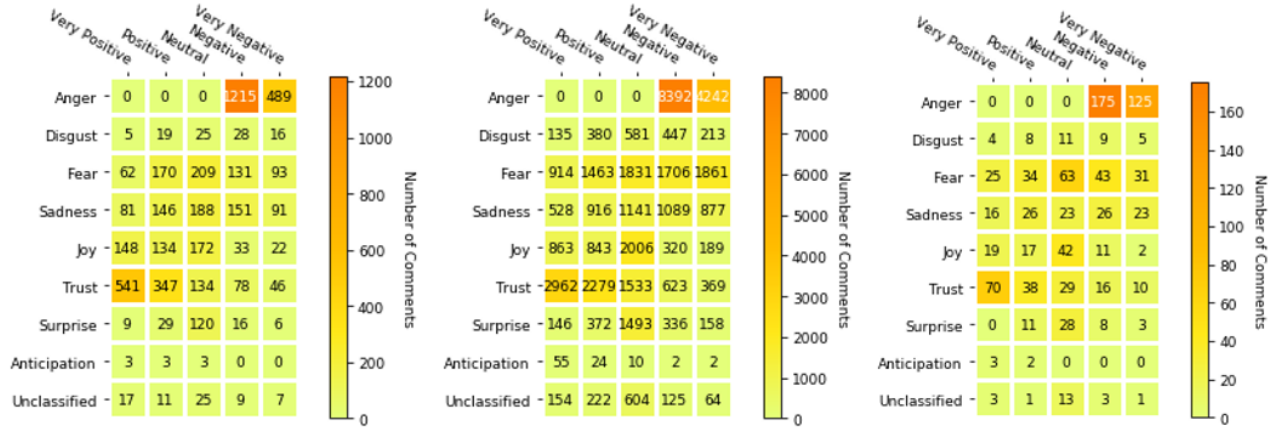
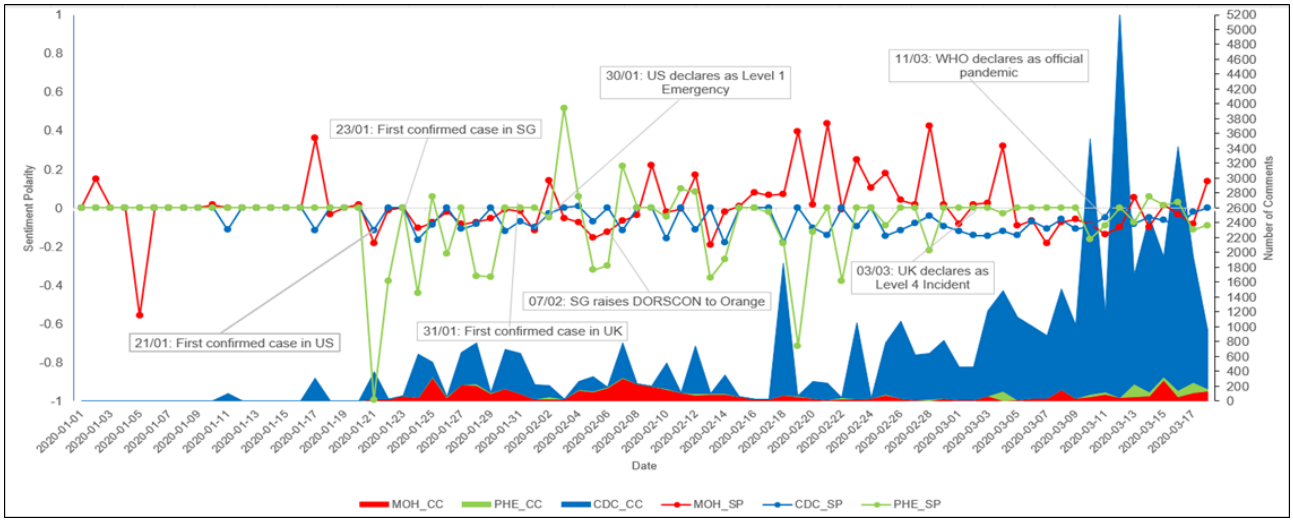


Figure 3. Temporal trend analysis for the number of comments and sentiment polarity. CDC_CC: Centers for Disease Control and Prevention's number of comments; CDC_SP: Centers for Disease Control and Prevention's sentiment polarity score; DORSCON: Disease Outbreak Response System Condition; MOH_CC: Ministry of Health's number of comments; MOH_SP: Ministry of Health's sentiment polarity score; PHE_CC: Pubilc Health England's number of comments; PHE_SP: Public Health England's sentiment polarity score; SG: Singapore; WHO: World Health Organization.



Identification of Toxicity in Comments of COVID-19 Posts

The CDC had the highest number of toxic comments, followed by the MOH and PHE (Table 5). PHE had the highest average of likes per toxic comment (LPTC) compared to the CDC and the MOH. For the PHE page, toxic comments received more likes and replies from its followers as compared to nontoxic comments. The mean LPTC for PHE was two times higher compared to nontoxic comments. Similarly, toxic comments received more replies compared to nontoxic comments.

Table 5. Summary of coronavirus disease toxic comments.

Agency	Total comments, n	Toxic comments, n (%)	LPTC[a], mean	LPNC[b], mean	RPTC[c], mean	RPNC[d], mean
MOH[e]	5032	58 (0.01)	0.47	1.38	0.14	0.43
CDC[f]	42,470	728 (0.01)	0.89	1.99	0.13	0.71
PHE[g]	977	12 (0.01)	2.42	1.44	0.92	0.60

[a]LPTC: likes per toxic comment.

[b]LPNC: likes per nontoxic comment.

[c]RPTC: replies per toxic comment.

[d]RPNC: replies per nontoxic comment.

[e]MOH: Ministry of Health.

[f]CDC: Centers for Disease Control and Prevention.

[g]PHE: Public Health England.

Discussion

Principal Findings

Among the three agencies, the MOH was the most active in using Facebook to reach out to its followers in terms of posting frequency, with an average of around 4 posts daily, exceeding both the CDC and PHE. The MOH displayed a similar active outreach strategy during the previous Zika outbreak in 2016 by ramping up engagement online with more frequent posting [21]. The MOH COVID-19 posts received more attention from their followers compared to the previous outbreak. For instance, the average number of comments received per post was 3.6 for Zika posts [21], while for COVID-19 posts, the average has increased to 15.6.

The posting frequency of the CDC and PHE on matters pertaining to COVID-19 was low in the initial peri-COVID-19 phase; this was because a substantial number of posts were still dedicated to other public health topics (eg, mental health, food disease outbreak, chronic disease management). This may reflect that both countries perceived the risk to be low or possibly that the outbreak was still largely confined within Asia. At the time of analysis, the epicenter of the COVID-19 pandemic was Wuhan, China with several other Asian countries, including Singapore. We observed that the volume of updates on COVID-19 related to prevention from the CDC and PHE increased toward the end of the analysis period, which parallels the surge in case count in both the United States and England. Given the large number of followers on the CDC and PHE Facebook pages, it is a missed opportunity that the CDC and PHE did not engage with their followers more intensively using Facebook. We observed that Facebook users who engaged with the CDC or PHE posts were more likely to share rather than to react or comment on the posts. Hence, Facebook may possibly be a useful platform for the CDC or PHE to disseminate information for Facebook users to propagate to others.

In our analysis of the PHAs' post content, we restricted the number of themes to six to focus on major themes, unlike earlier studies where more themes were used with a meagre number of posts for certain themes [26]. The posts from the MOH were more diverse with frequent updates on preventive measures, travel advisories, disease information, falsehood correction, and even appreciation for health care workers and other frontline staff. This was in contrast to the posts from the CDC and PHE, where the messages were mostly focused on preventive measures and sporadic situation updates. Falsehood corrections are in need during this pandemic, as an earlier study identified that misleading Facebook posts acquired more popularity than accurate posts during the Zika outbreak in the United States [43]. On this point, we observed that the MOH has adopted misinformation debunking as one of its community and social measures for handling the COVID-19 situation [6]. The current distribution of themes may reflect the different phases that the three countries were going through during the time of our analyses. We anticipate that as the pandemic develops across each country the themes of the posts will continually evolve. Nevertheless, there is a need to enhance awareness and not undermine the possibility of a serious outbreak during the precrisis period [22]. From our analyses, we did not identify much evidence to show attempts at such efforts from the CDC and PHE in their official Facebook pages.

In previous disease pandemics, negative sentiments were generally prevalent in social media [23]. We made similar findings in our analyses where the majority of the posts conveyed anger emotions and negative sentiments. For the MOH, however, we noticed that over time (from mid-February to the first week of March 2020), Facebook users began to be more positive about the government's response to the pandemic. This demonstrates that monitoring sentiments and emotions on social media can help PHAs gauge the effectiveness of their public health education efforts on Facebook. Another observation that supports the monitoring of sentiments and emotions on social media is that the number of comments tend

to spike in conjunction with specific events. For example, our data showed that the number of comments sharply rose in association with the first confirmed case in the United States (January 21, 2020), when Singapore rose its DORSCON level to orange (February 7, 2020), when the United Kingdom declared COVID-19 as a Level 4 incident (March 3, 2020), and when the WHO declared COVID-19 a pandemic (March 11, 2020).

The prevalence of toxic comments for all three Facebook pages' COVID-19 posts was fairly low. It is possible that the majority of the toxic comments had been removed (for instance, the CDC has a policy that profane and obscene comments can be deleted [44]), and what was analyzed were those that were not filtered. The US President has repeatedly referred to SARS coronavirus 2 as the "Chinese virus," and this may have led to anti-Asian sentiments [45]. This may explain why the CDC had the highest number of toxic comments. Fortunately, the volume of such comments remained low, and the agreement with such toxic comments was also low as reflected by the low average number of likes. PHAs should consider dedicating resources during a pandemic to manage toxic comments as well as combat falsehood. We observed that both the CDC and PHE did not have any post to correct falsehoods, unlike the MOH. In contrast, social media platforms have been proactive in setting up centralized hubs dedicated to COVID-19 updates and information such as the COVID-19 Information Center on Facebook [46] or the COVID-19 Information and Resources collection on Google [47] to direct social media users to trusted and reliable information.

Limitations

In this study, we analyzed data from Facebook only. However, PHAs may have used other social media platforms such as Twitter and YouTube to disseminate public health information to their citizens. Thus, this study's findings may not fully represent the overall social media outreach efforts of PHAs during the COVID-19 pandemic. In addition, the sentiments captured on Facebook comments may not reflect the users of other social media platforms, as the user profiles of these various platforms are known to be different. Furthermore, PHAs may currently still use traditional news and mass media channels to reach the public with information, updates, and guidance measures. Hence, PHAs' outreach efforts in social media platforms is supplementary, and these platforms are considered either as a resource for additional information or for reaching out to people who no longer follow traditional news and mass media channels. Another limitation of the study was that we have limited our analyses of the Facebook followers' response to posts initiated by the PHAs. We have not, for instance, analyzed the comments within private circles or closed groups that may be different in nature compared to publicly disclosed comments. Our paper has focused on the comparative analysis of how three PHAs have used Facebook for COVID-19 communications strategy. We did not evaluate if the PHAs' use of social media has improved over time as they gain experience from dealing with other infectious diseases such as the H1N1 swine flu pandemic, Ebola epidemic, or Zika outbreak. As social media platforms become more prominent, its users and the interactions among its users evolve. Optimal use of these platforms for public health communications will benefit from constant reflection and critical appraisal of what strategies have worked and what have not. Finally, the thematic analysis may be more robust if independent coding of the posts was conducted. However, the huge number of posts rendered this process time-consuming. Accordingly, the review and confirmation of the themes from a second coder was sought as an acceptable compromise.

Conclusions

The Facebook postings by the PHAs in this study provided some insights into their governments' COVID-19 broader communication strategy. Through our study, we identified differences in the Facebook-based outreach and engagement efforts of three developed countries during the prepandemic and peripandemic periods of COVID-19. The differences were found in terms of both posting frequency and themes in posts. The change in sentiments in response to specific outreach events were also observed. On the whole, the MOH stepped up its outreach efforts on Facebook more intensively compared to the CDC and PHE. We hope that our findings will be of interest to PHAs and health science researchers who study pandemics in the context of social media. In our upcoming work, we intend to conduct studies in two related directions. In our first set of studies, we intend to collect more data from the same three PHA Facebook pages and analyze the results with data segregated in three phases (pre-COVID-19, peri-COVID-19, and post-COVID-19). In the second set of studies, we intend to analyze the outreach efforts of other countries during this COVID-19 pandemic to understand the effectiveness and shortfalls of strategies used by different countries.

Acknowledgments

We would like to thank Miss Yu-Ting, CHEN (research assistant) for providing administrative support to submitting this manuscript. ASR was hired under a grant from the National University Health System Strategic Funds titled "Phen-Gen Research Repository and Pre-Emptive Pharmacogenomics." During COVID-19, safe distancing measures prevented him from accessing data stored in a dedicated workstation without internet access. He was thus reassigned to this study. The grantor had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication. This work was also supported by the Centre for Health Services and Policy Research, a center under the National University Health Systems Pte Ltd and Saw Swee Hock School of Public Health, National University of Singapore. We would like to thank the Perspective API team of Google for generously extending the usage quota of their API service, which we used to measure the toxicity of Facebook comments for our research study.

Authors' Contributions

H-LW conceptualized the study, interpreted the data, critically revised the manuscript for important intellectual content, and provided supervision. ASR designed the study; acquired, analyzed, and interpreted the data; drafted the manuscript; and critically revised the manuscript for important intellectual content. SGT analyzed and interpreted the data and critically revised the manuscript for important intellectual content. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Emojis and the associated emotions.
[DOCX File , 16 KB - jmir_v22i5e19334_app1.docx]

References

1. Worldometer. 2020. Countries where COVID-19 has spread URL: https://www.worldometers.info/coronavirus/countries-where-coronavirus-has-spread/ [accessed 2020-04-11]
2. Uras U. Al Jazeera. 2020 Apr 08. Coronavirus: comparing COVID-19, SARS and MERS URL: https://www.aljazeera.com/news/2020/04/coronavirus-comparing-covid-19-sars-mers-200406165555715.html [accessed 2020-04-11]
3. Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, et al. First case of 2019 novel coronavirus in the United States. N Engl J Med 2020 Mar 05;382(10):929-936. [doi: 10.1056/nejmoa2001191]
4. Moss P, Barlow G, Easom N, Lillie P, Samson A. Lessons for managing high-consequence infections from first COVID-19 cases in the UK. Lancet 2020 Mar 14;395(10227):e46 [FREE Full text] [doi: 10.1016/S0140-6736(20)30463-3] [Medline: 32113507]
5. Abdullah Z, Salamat H. Channel News Asia. 2020 Jan 23. Singapore confirms first case of Wuhan virus URL: https://www.channelnewsasia.com/news/singapore/wuhan-virus-pneumonia-singapore-confirms-first-case-12312860 [accessed 2020-04-11]
6. Lee V, Chiew C, Khong W. Interrupting transmission of COVID-19: lessons from containment efforts in Singapore. J Travel Med 2020 Mar 13:e [FREE Full text] [doi: 10.1093/jtm/taaa039] [Medline: 32167146]
7. Ministry of Health. 2020 Feb 07. Risk assessment raised to DORSCON orange URL: https://www.moh.gov.sg/news-highlights/details/risk-assessment-raised-to-dorscon-orange [accessed 2020-04-11]
8. World Health Organization. 2020 Mar 11. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020 URL: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2020-04-11]
9. Barron L. Time. 2020 Mar 13. What we can learn from Singapore, Taiwan and Hong Kong about handling coronavirus URL: https://time.com/5802293/coronavirus-covid19-singapore-hong-kong-taiwan/ [accessed 2020-04-10]
10. Teo J. The Straits Times. 2020 Mar 10. Coronavirus: WHO praises Singapore's containment of Covid-19 outbreak URL: https://www.straitstimes.com/singapore/health/coronavirus-who-praises-singapores-containment-of-covid-19-outbreak [accessed 2020-04-09]
11. The Straits Times. 2020 Mar 13. Covid-19 situation is serious but under control in Singapore: PM Lee URL: https://www.straitstimes.com/singapore/covid-19-situation-is-serious-but-under-control-in-spore-pm-lee [accessed 2020-04-09]
12. Facebook. Facebook. 2020 Mar 31. Coronavirus: lessons learned from Singapore URL: https://www.facebook.com/nus.sshsph/posts/2766847946771712?__tn__=-R [accessed 2020-04-09]
13. Facebook. 2020 Apr 08. COVID-19: lessons learned? What comes next? URL: https://www.facebook.com/nus.sshsph/posts/2784851854971321 [accessed 2020-04-09]
14. Costello A. The Guardian. 2020 Mar 15. The UK's Covid-19 strategy dangerously leaves too many questions unanswered URL: https://www.theguardian.com/commentisfree/2020/mar/15/uk-covid-19-strategy-questions-unanswered-coronavirus-outbreak [accessed 2020-04-10]
15. Biesecker M. Global News. 2020 Apr 05. U.S. 'wasted' months before preparing for COVID-19 pandemic: former health secretary URL: https://globalnews.ca/news/6780562/coronavirus-covid-wasted-months/ [accessed 2020-04-10]
16. Hoffman SJ, Silverberg SL. Delays in global disease outbreak responses: lessons from H1N1, Ebola, and Zika. Am J Public Health 2018 Mar;108(3):329-333. [doi: 10.2105/AJPH.2017.304245] [Medline: 29345996]
17. Holmes BJ. Communicating about emerging infectious disease: the importance of research. Health Risk Soc 2008 Aug;10(4):349-360. [doi: 10.1080/13698570802166431]
18. Vaughan E, Tinker T. Effective health risk communication about pandemic influenza for vulnerable populations. Am J Public Health 2009 Oct;99(S2):S324-S332. [doi: 10.2105/ajph.2009.162537]
19. Wong JEL, Leo YS, Tan CC. COVID-19 in Singapore-current experience: critical global issues that require attention and action. JAMA 2020 Feb 20:e. [doi: 10.1001/jama.2020.2467] [Medline: 32077901]
20. Vijaykumar S, Jin Y, Nowak G. Social media and the virality of risk: the risk amplification through media spread (RAMS) model. J Homeland Secur Emerg Manage 2015:653. [doi: 10.1515/jhsem-2014-0072]

21. Vijaykumar S, Meurzec R, Jayasundar K, Pagliari C, Fernandopulle Y. What's buzzing on your feed? Health authorities' use of Facebook to combat Zika in Singapore. J Am Med Informatics Assoc 2017;24(6):1159. [doi: 10.1093/jamia/ocx028]
22. Lwin M, Lu J, Sheldenkar A, Schulz P. Strategic uses of Facebook in Zika outbreak communication: implications for the crisis and emergency risk communication model. Int J Environ Res Public Health 2018 Sep 10;15(9):1974. [doi: 10.3390/ijerph15091974] [Medline: 30201929]
23. Mamidi R, Miller M, Banerjee T, Romine W, Sheth A. Identifying key topics bearing negative sentiment on Twitter: insights concerning the 2015-2016 Zika epidemic. JMIR Public Health Surveill 2019 Jun 04;5(2):e11036. [doi: 10.2196/11036] [Medline: 31165711]
24. de Araujo DHM, de Carvalho EA, da Motta CLR, da Silva Borges MR, Gomes JO, de Carvalho PVR. Social networks applied to Zika and H1N1 epidemics: a systematic review. Springer, Cham 2019:679. [doi: 10.1007/978-3-319-96089-0_74]
25. Strekalova YA. Health risk information engagement and amplification on social media. Health Educ Behav 2017 Apr;44(2):332-339. [doi: 10.1177/1090198116660310] [Medline: 27413028]
26. Vijaykumar S, Nowak G, Himelboim I, Jin Y. Virtual Zika transmission after the first U.S. case: who said what and how it spread on Twitter. Am J Infect Control 2018 May;46(5):549-557. [doi: 10.1016/j.ajic.2017.10.015] [Medline: 29306490]
27. World Health Organization. 2018 Jan 10. Communicating risk in public health emergencies: a WHO guideline for emergency risk communication (ERC) policy and practice URL: https://www.who.int/risk-communication/guidance/download/en/ [accessed 2020-04-09]
28. Glik DC. Risk communication for public health emergencies. Annu Rev Public Health 2007;28:33-54. [doi: 10.1146/annurev.publhealth.28.021406.144123] [Medline: 17222081]
29. Jünger J, Keyling T. GitHub. 2019. Facepager URL: https://github.com/strohne/Facepager/ [accessed 2020-04-06]
30. Facebook. Ministry of Health, Singapore URL: https://www.facebook.com/sghealthministry [accessed 2020-05-08]
31. Facebook. CDC URL: https://www.facebook.com/CDC [accessed 2020-05-08]
32. Facebook. Public Health England URL: https://www.facebook.com/PublicHealthEngland/ [accessed 2020-05-08]
33. Ministry of Health. 2020 Jan 02. Precautionary measures in response to severe pneumonia cases in Wuhan, China URL: https://www.moh.gov.sg/news-highlights/details/precautionary-measures-in-response-to-severe-pneumonia-cases-in-wuhan-china [accessed 2020-04-10]
34. Erlingsson C, Brysiewicz P. A hands-on guide to doing content analysis. Afr J Emerg Med 2017 Sep;7(3):93-99. [doi: 10.1016/J.AFJEM.2017.08.001]
35. Hutto C, Gilbert E. VADER: a parsimonious rule-based model for sentiment analysis of social media text. AAI Publications, Eighth International AAAI Conference on Weblogs and Social Media 2014:A [FREE Full text]
36. Plutchik R. Emotions: a general psychoevolutionary theory. In: Approaches to Emotion. London: Psychology Press; May 22, 2014:A.
37. Felbo B, Mislove A, Søgaard A, Rahwan I, Lehmann S. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. 2017 Presented at: 2017 Conference on Empirical Methods in Natural Language Processing; September 2017; Copenhagen, Denmark. [doi: 10.18653/v1/d17-1169]
38. Evers C, Adriaanse M, de Ridder DT, de Witt Huberts JC. Good mood food. Positive emotion as a neglected trigger for food intake. Appetite 2013 Sep;68:1-7. [doi: 10.1016/j.appet.2013.04.007] [Medline: 23602962]
39. Al-Hajjar D, Syed A. Applying sentiment and emotion analysis on brand tweets for digital marketing. 2015 Dec 21 Presented at: 2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT); 3-5 Nov 2015; Amman, Jordan. [doi: 10.1109/aeect.2015.7360592]
40. Daniulaityte R, Lamy F, Smith G, Nahhas R, Carlson R, Thirunarayan K, et al. "Retweet to pass the blunt": analyzing geographic and content features of cannabis-related tweeting across the United States. J Stud Alcohol Drugs 2017 Nov;78(6):910-915 [FREE Full text] [doi: 10.15288/jsad.2017.78.910] [Medline: 29087826]
41. Cheng J, Bernstein M, Danescu-Niculescu-Mizil C, Leskovec J. Anyone can become a troll: causes of trolling behavior in online discussions. 2017 Presented at: 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing; 2017; New York, NY p. 1217-1230. [doi: 10.1145/2998181.2998213]
42. Google. Perspective API URL: https://www.perspectiveapi.com/#/home [accessed 2020-04-06]
43. Sharma M, Yadav K, Yadav N, Ferdinand KC. Zika virus pandemic-analysis of Facebook as a social media health information platform. Am J Infect Control 2017 Mar 01;45(3):301-302. [doi: 10.1016/j.ajic.2016.08.022] [Medline: 27776823]
44. CDC. Social media public comment policy URL: https://www.cdc.gov/socialmedia/tools/CommentPolicy.html [accessed 2020-05-08]
45. The Washington Post. 2020. Trump calling coronavirus "Chinese virus" encourages racism against Asian Americans, experts say URL: https://www.washingtonpost.com/nation/2020/03/20/coronavirus-trump-chinese-virus/ [accessed 2020-04-09]
46. Facebook. Coronavirus (COVID-19) Information Centre URL: https://www.facebook.com/coronavirus_info/ [accessed 2020-04-10]
47. Google. COVID-19 information & resources URL: https://www.google.com/covid19/ [accessed 2020-04-13]

Abbreviations

API: application program interface
CDC: Centers for Disease Control and Prevention
COVID-19: coronavirus disease
CPP: comments per post
DORSCON: Disease Outbreak Response System Condition
LPTC: likes per toxic comment
MOH: Ministry of Health
PHA: public health authority
PHE: Public Health England
PNER: positive to negative emotions ratio
PNSR: positive to negative sentiment ratio
PPD: posts per day
RPP: reactions per post
SARS: severe acute respiratory syndrome
SP: sentiment polarity
SPP: shares per post
WHO: World Health Organization

Edited by G Eysenbach; submitted 13.04.20; peer-reviewed by G Nowak, A Dormanesh, A Natale, C Mather; comments to author 06.05.20; revised version received 11.05.20; accepted 12.05.20; published 19.05.20.

Please cite as:
Sesagiri Raamkumar A, Tan SG, Wee HL
Measuring the Outreach Efforts of Public Health Authorities and the Public Response on Facebook During the COVID-19 Pandemic in Early 2020: Cross-Country Comparison
J Med Internet Res 2020;22(5):e19334
URL: http://www.jmir.org/2020/5/e19334/
doi:10.2196/19334
PMID:32401219

Original Paper

Telehealth Home Support During COVID-19 Confinement for Community-Dwelling Older Adults With Mild Cognitive Impairment or Mild Dementia: Survey Study

Jessica Marian Goodman-Casanova[1], RNCS; Elena Dura-Perez[1], MSc; Jose Guzman-Parra[1], PhD; Antonio Cuesta-Vargas[2], PhD; Fermin Mayoral-Cleries[1], PhD

[1]Department of Mental Health, Regional University Hospital of Málaga, Biomedical Research Institute of Malaga (IBIMA), Málaga, Spain

[2]Department of Physiotherapy, University of Málaga, Biomedical Research Institute of Malaga (IBIMA), Málaga, Spain

Corresponding Author:
Jessica Marian Goodman-Casanova, RNCS
Department of Mental Health
Regional University Hospital of Málaga
Biomedical Research Institute of Malaga (IBIMA)
Plaza del Hospital Civil s/n
Málaga, 29009
Spain
Phone: 34 660901966
Email: jmariangoodman@gmail.com

Abstract

Background: The public health emergency of coronavirus disease (COVID-19) is rapidly evolving worldwide; some countries, including Spain, have implemented restrictive measures. Populations that are vulnerable to this outbreak and its physical and mental health effects include community-dwelling older adults with mild cognitive impairment or mild dementia. Telehealth is a potential tool to deliver health care and decrease exposure risk.

Objective: The aims of this study were to explore the impact of confinement on the health and well-being of community-dwelling older adults with mild cognitive impairment or mild dementia, to provide television-based and telephone-based health and social support, and to study the effects of a television-based assistive integrated technology, TV-AssistDem (TeleVision-based ASSistive Integrated Service to supporT European adults living with mild DEMentia or mild cognitive impairment).

Methods: A telephone-based survey was administered in Spain to 93 participants in the TV-AssistDem clinical trial from March 25 to April 6, 2020.

Results: Of the respondents, 60/93 (65%) were women. The mean age was 73.34 (SD 6.07), and 69/93 (74%) lived accompanied. Lockdown measures forced 17/93 respondents (18%) to change their living arrangements. Health status was found to be optimal in 89/93 respondents (96%), with no COVID-19 symptoms. Grocery and pharmacy outings were performed by family members of 68/93 participants (73%); 57 (61%) reported overall well-being, and 65 (70%) maintained their sleep quality. However, participants living alone reported greater negative feelings and more sleeping problems. Regarding leisure activities, 53/93 respondents (57%) took walks, 32 (35%) played memory games, 55 (60%) watched television, and 91 (98%) telephoned relatives. 58/93 (64%) respondents reported accessing moderate or too much COVID-19 information, 89 (97%) received it from television, and 56 (62%) stated that their understanding of the information was extreme. 39/93 (39%) respondents had contacted health and social services, while 29 (31%) requested information regarding these services during the telephone call. There were no significant differences in health and well-being between the intervention and control groups. Respondents with TV-AssistDem performed more memory exercises (24/93, 52% vs 8/93, 17.4%; $P<.001$) than control respondents.

Conclusions: Our findings suggest that during COVID-19 confinement, the physical and mental health and well-being was optimal for the majority of our vulnerable population. However, those living alone reported greater negative psychological effects and sleeping problems. Measures adopted to address the negative experiences of confinement included keeping informed about the situation, accessing health and social services, having a support network that prevents risk of exposure to COVID-19 and guarantees food and medical supplies, a daily routine with maintained sleeping habits and leisure activities, staying physically and mentally active with cognitive stimulation exercises, and ensuring social connectedness using technology. Television sets were preferred technological devices to access COVID-19 information, watch television as a recreational activity, and perform

memory exercises as an intellectual activity. Television-based telehealth support using TV-AssistDem demonstrated potential for cognitive stimulation.

Trial Registration: ClinicalTrials.gov NCT03653234; https://clinicaltrials.gov/ct2/show/NCT03653234

(J Med Internet Res 2020;22(5):e19434) doi:10.2196/19434

KEYWORDS

telehealth; confinement; coronavirus; COVID-19; well-being; emergency response; public health; mental health; physical health; elderly; older adults; cognitive impairment; dementia

Introduction

Coronavirus disease (COVID-19) has been declared a global emergency by the World Health Organization (WHO); this disease has created a rapidly evolving situation which has forced the implementation of unprecedented restrictions to control its viral spread and mitigate its impact [1]. Spain has one of the highest burdens of COVID-19 worldwide, with 59.4% of cases reported in people aged 60 years and older [2]. In response to the outbreak, the Spanish government issued a royal decree (463/2020) to declare a 15-day national emergency, with the exceptional measure of a nationwide lockdown that started on March 15th and has been extended twice since [3]. This decree enforces social distancing, quarantine of people exposed to the disease, and home confinement of people who remain healthy, allowing only essential outings. This restriction of movement of daily life activities and separation from loved ones may be challenging and unpleasant. This experience may impact the physical and mental health and well-being of those who undergo it. Demographic factors, such as sex, age, and baseline health and well-being status, have been described as preconfinement predictors of greater impact [4].

Community-dwelling older adults are among the most vulnerable to the impact of this confinement. Their chronic conditions may be aggravated by the consequences of confinement [4]. Additionally, WHO underlines that people with mild cognitive impairment or mild dementia may face a greater challenge during the outbreak [5]. Their comprehension of the public health situation and the measures to follow, such as staying at home or wearing a mask, may be limited [6]. Confinement may lead to social isolation, which is a risk factor for health-related consequences [7] and increases risk of dementia and cognitive decline in older adults [8-10]. Moreover, facing a novel and unknown situation is a potential stressor, especially when cognition is compromised [4]. Furthermore, nonattendance of face-to-face activities, such as memory workshops and day care services, may worsen the cognition and functioning of this population [11]. Reduced support availability may consequently increase caregiver burden [6]. Alzheimer Europe recommends the following for people with mild cognitive impairment or mild dementia and their caregivers: building a support network; keeping well informed; guaranteeing food and medical supplies; enjoying leisure activities; staying physically and mentally active using Stimulus, among others, for cognitive stimulation; and keeping socially connected [12].

In the age of information and communications technology, technology home-based interventions [13] (smartphones, tablets, computers, smart televisions, virtual assistants, ambient assistive devices, etc.) can support most of the above recommendations; they facilitate information sharing and online shopping, provide access to sports and entertainment, increase social connectedness, etc. However, most importantly, these interventions enable the distribution of health-related information and services. Telehealth is a potential major tool to deliver routine health care and prevent the risk of viral exposure, especially for people at higher risk [14].

The burden of COVID-19 exerts pressure on health care services, social support services [3], and caregivers. Public health systems that had not proactively integrated telehealth are working reactively against the clock to respond to this urgent situation. Nationally, the Spanish government has launched a COVID-19 triage app, Asistencia COVID-19 [15]; regionally, routine care is being remotely managed via centralized telephone numbers (Salud Responde) [16]. These efforts are noteworthy; however, for years prior to this outbreak, Europe proactively invested in information and communications technology research in at-risk populations with programs such as the European Active and Assisted Living (AAL) Programme. TV-AssistDem (TeleVision-based ASSistive Integrated Service to supporT European adults living with mild DEMentia or mild cognitive impairment) is an AAL project that was selected for the Call for Proposals “Living well with dementia: The contribution of information and communications technology to integrated solutions for enabling the well-being of people living with dementia and their communities” [17].

TV-AssistDem is a European multicenter randomized controlled trial that evaluates a television-based assistive integrated service to support and improve the quality of life of people with mild cognitive impairment or mild dementia and provide relief to their caregivers. The components of TV-AssistDem are a digital set-top-box service based on Android technology with a television-based interface, a webcam, and a centralized back-end service with a web-based interface. TV-AssistDem facilitates remote support through data transmission and video interactivity between users, caregivers, and health care professionals.

The current study is nested in the TV-AssistDem project. To address the unexpected health emergency of COVID-19, the TV-AssistDem team rapidly adapted the service to provide tailored support. Detailed information on COVID-19 was offered through the functionality of health education with selected content from official sources, such as the WHO, the Spanish Ministry of Health, and local authorities. Videos on recommendations and basic care measures, such as hand washing, were uploaded. In addition to offering informational content, three of the established functionalities of TV-AssistDem

guarantee physical and mental health and well-being, social connectedness, and cognitive stimulation. The Health Education functionality enables visualization of videos of physical activity at home; the Videocall functionality enables communication with loved ones and health professionals through videocalls; and the Memory Games functionality provides cognitive stimulation with Stimulus memory games [18].

The use of information and communications technology in reducing social isolation, improving cognition, and facilitating access to services in people with mild cognitive impairment or mild dementia has been broadly studied [13,19]. However, despite the growing global interest in telehealth during the COVID-19 pandemic [14,20], no study has yet explored the use of telehealth home support during COVID-19 confinement in people with mild cognitive impairment or mild dementia and the impact of confinement on this population.

The three aims of this study were to explore the impact of confinement on the physical and mental health and well-being of community-dwelling older adults with mild cognitive impairment or mild dementia, to provide television-based and telephone-based health and social support, and to study the effects of a television-based assistive integrated technology (TV-AssistDem). We hypothesized that people with access to TV-AssistDem would report greater physical and mental health and well-being.

Methods

Ethical Declarations

The current study is nested in the clinical trial TV-AssistDem (ClinicalTrials.gov NCT03653234) approved by the Malaga Province Research Ethics Committee (Comité de Ética de la Investigación Provincial de Málaga), approval number 1770-N-17. The substantial amendments derived from this new study were reviewed by the ethics committee and were granted a favorable opinion.

People with mild cognitive impairment or mild dementia and their caregivers provided written consent before taking part in the TV-AssistDem clinical trial as per protocol [18]. Taking into consideration the rights of the participants and to perform this study during this exceptional situation of confinement, we informed people with mild cognitive impairment or mild dementia and their caregivers at the beginning of the telephone interview that the reason for the call was additional follow-up within the framework of the project.

Survey Development

Telephone-based survey research was conducted according to Gordon's Functional Health Patterns [21] (Multimedia Appendix 1). Overall, quantitative strategies (questions with numerically rated items) were used for data collection of the health perception-health management and sleep-rest patterns. Qualitative strategies (open-ended questions) were used for the coping-stress tolerance, activity-exercise, and role-relationship patterns. Data were organized into previously coded and listed categories using a directed content analysis approach.

Participant Identification and Recruitment

Researchers from the Biomedical Research Institute of Malaga contacted 100 community-dwelling older adults with mild cognitive impairment or mild dementia by telephone. Potential respondents were TV-AssistDem study participants from both the intervention and control groups who had met eligibility criteria for the TV-AssistDem trial and who had not dropped out of the study. Participants in the intervention group had been specifically trained in the use of TV-AssistDem and were given daily access to the service in their home environment. Participants in the control group received treatment as usual. Both groups received follow-up visits at 6 and 12 months. Inclusion criteria were age >60 years, self-perceived cognitive impairment or caregiver's perception of cognitive impairment that was present for at least 6 months, score of 23-27 points on the Mini-Mental State Examination, independent living, informal caregiver, pharmacological treatment, and written consent. Exclusion criteria were a score >11 on the Geriatric Depression Scale, terminal illness, and specific conditions (cognitive, visual, motor, etc.) which could compromise the use of the system [18].

Interview Process

Participants were contacted by telephone by health professionals (a mental health registered nurse clinical specialist and a neuropsychologist). The researchers had previously established relationships with the participants during the TV-AssistDem study. Potential respondents were considered unreachable when no answer was given to 3 different calls on 3 different days. The telephone call time frame was March 25 to April 6, 2020. The researchers interviewed the participants using the telephone-based survey. Caregivers were interviewed on behalf of people with mild cognitive impairment or mild dementia when cognitive or emotional statuses were compromised.

During each interview, when necessary, health information and counselling was provided by the health professionals regarding COVID-19. This information included the symptoms and mode of transmission of the disease, contact telephone numbers for health care and social services to manage difficulties arising from the confinement situation, and recommendation guidelines for staying physically and mentally active. In addition, people with mild cognitive impairment or mild dementia in the intervention group were notified of the new updates available in the Health Education functionality TV-AssistDem concerning COVID-19. These included selected content from official sources such as the WHO, the Spanish Ministry of Health, and local authorities. Informative content included infographics and videos about the disease, basic protection measures such as hand washing, advice for managing psychological distress, and guidelines to carry out health procedures. Additionally, the participants were encouraged to continue to use three of the usual functionalities of TV-AssistDem which are meant to guarantee physical and mental health and well-being. Physical activity was promoted through the visualization of videos of indoor home exercise. Cognition was stimulated with Stimulus memory games. Lastly, social connectedness with loved ones and health professionals was facilitated through videocalls (Figures 1-3).

Figure 1. TV-AssistDem adapted to provide tailored support during COVID-19.

Figure 2. TV-AssistDem focused on cognitive stimulation and COVID-19 information.

Figure 3. TV-AssistDem focused on social connectedness and physical activity.

Data Analysis

The chi-square test and Fisher exact test (when fewer than 80% of the expected frequencies of the cell were greater than 5) were used for analysis of the categorical variables. For quantitative variables, the Student *t* test was used. The analysis was performed by following an intention-to-treat procedure. The R program (version 3.6.2) was used for the analysis.

Results

Participants

A total of 93/100 (93.0%) TV-AssistDem study participants were successfully contacted: 47/93 (51%) in the intervention group and 46/93 (49%) in the control group. We could not reach 7/100 (7.0%) of the participants. The mean duration of the telephone calls was 12 minutes and 4 seconds (SD 7 seconds). For 21/93 (22%) participants, caregivers were interviewed on behalf of people with mild cognitive impairment or mild dementia whose cognitive or emotional statuses were compromised.

Sociodemographics

The sample of people with mild cognitive impairment or mild dementia had a mean age of 73.34 years (SD 6.07); 60/93 (65%) were women, and 69 (74%) lived accompanied. Lockdown measures forced 17/93 participants (18%) to change their living arrangements (Table 1).

Table 1. Sample characteristics and differences between the intervention and control groups regarding demographic characteristics and living arrangements.

Characteristic	Total (N=93)	Intervention (n=47)	Control (n=46)	Statistical difference	*P* value
Age (years), mean (SD)	73.34 (6.07)	74.00 (6.16)	72.67 (5.98)	t_{91}=1.053	.29
Sex, n (%)					
Male	33 (36)	16 (34)	17 (37)	χ^2_1=0.09	.77
Female	60 (65)	31 (66)	29 (33)		
Change in living arrangements due to lockdown, n (%)					
Yes	17 (18)	10 (21)	7 (15)	χ^2_1=0.57	.45
No	76 (82)	37 (79)	39 (85)		
Living arrangements, n (%)					
Alone	24 (26)	14 (30)	10 (22)	χ^2_4=3.75	.44
Spouse	39 (42)	22 (47)	17 (37)		
Children	12 (13)	5 (11)	7 (15)		
Spouse and children	13 (14)	4 (9)	9 (19)		
Other	5 (6)	2 (4)	3 (7)		

Functional Health Patterns

The health status of the participants was mainly found to be optimal; 89/93 (97%) presented no COVID-19 symptoms. Grocery and pharmacy outings were performed by family members for 68/93 participants (73%). Most of our respondents did not report inadequate or insufficient food supplies, nor did they report being unable to obtain regular medical care or prescriptions. 57/93 (61%) respondents reported overall well-being, and 65 (70%) maintained their quality of sleep (Table 2). Negative experiences reported included fear of becoming infected or infecting family members, frustration and boredom due to not being able to take part in daily activities, loss of usual routine, and social isolation. Leisure activities included physical, intellectual, recreational, and social activities. Of the 93 participants, 53 (57%) took walks, 32 (35%) played memory games, 55 (59%) watched TV, and 91 (98%) telephoned their family and friends (Table 3). Off-protocol, numerous respondents mentioned during the interviews that religious activities such as listening to religious programs on the radio felt comforting, and joining their neighbors at 8 PM for the national clapping against COVID-19 felt uplifting and provided a powerful reminder that although restrictions are followed in isolation, we are all part of a community.

Table 2. Differences between the intervention and control groups regarding health status and management during the COVID-19 pandemic.

Health status	Total (N=93), n (%)	Intervention (n=47), n (%)	Control (n=46), n (%)	Chi-square (*df*)	*P* value
Health status (COVID-19)[a]					
No symptoms	89 (96)	45 (96)	44 (96)	1.33 (2)	.51
Symptoms without test	3 (3)	1 (2)	2 (4)		
Symptoms and positive test	0 (0)	0 (0)	0 (0)		
Hospitalized	0 (0)	0 (0)	0 (0)		
ICU[b] inpatient	1 (1)	1 (2)	0		
Deceased	0	0	0		
Health management: groceries/pharmacy					
Patient	12 (13)	3 (6)	9 (20)	9.19 (5)	.10
Patient and family member	7 (8)	3 (6)	4 (9)		
Family member	68 (73)	35 (75)	33 (72)		
Home worker	1 (1)	1 (2)	0 (0)		
Online	1 (1)	1 (2)	0 (0)		
Other	4 (4)	4 (9)	0 (0)		
Mental health and well-being					
Well	57(61)	27 (59)	30 (65)	0.41 (1)	.52
Calm	8 (9)	3 (7)	5 (11)	0.55 (1)	.46
Sad	27 (29)	17 (37)	10 (22)	2.57 (1)	.11
Worried	20 (22)	7 (15)	13 (28)	2.30 (1)	.13
Afraid	10 (11)	6 (13)	4 (9)	0.450 (1)	.50
Anxious	22 (24)	8 (17)	14 (30)	2.15 (1)	.14
Bored	13 (14)	6 (13)	7 (15)	0.09 (1)	.74
Sleep quality					
Maintained	65 (70)	35 (81)	30 (68)	2.01 (1)	.16
Altered	22 (24)	8 (19)	14 (32)		
Unknown	6 (6)	0 (0)	0 (0)		

[a]COVID-19: coronavirus disease.

[b]ICU: intensive care unit.

Table 3. Differences between the intervention and control groups regarding activities.

Activity category and type	Total (N=93), n (%)	Intervention (n=47), n (%)	Control (n=46), n (%)	Chi-square (*df*)	*P* value
Physical					
None	12 (13)	8 (19)	4 (9)	1.43 (1)	.23
Walking	53 (57)	23 (49)	30 (65)	2.51 (1)	.11
Stair climbing	10 (11)	5 (11)	5 (11)	0.01 (1)	.97
Gymnastics	19 (20)	12 (26)	7 (12)	1.52 (1)	.22
House chores	8 (9)	6 (13)	2 (4)	2.09 (1)	.27
Other	12 (13)	3 (6)	9 (20)	3.59 (1)	.058
Intellectual					
Memory exercises	32 (35)	24 (52)	8 (17)	12.22 (1)	<.001
Reading	24 (26)	13 (28)	11 (24)	0.22 (1)	.63
Playing games	6 (7)	1 (2)	5 (11)	2.85 (1)	.20
Needlework	18 (20)	6 (13)	12 (26)	2.49 (1)	.11
Painting	8 (9)	5 (11)	3 (7)	0.55 (1)	.71
Recreational					
Watching television	55 (60)	28 (61)	27 (59)	0.04 (1)	.83
Listening to radio or music	9 (9.8)	3 (6.5)	6 (13)	1.11 (1)	.48
Playing with information and communications technology	8 (8.7)	4 (8.7)	4 (9)	0.00 (1)	>.99
House chores	41 (45)	18 (39)	23 (50)	1.10 (1)	.29
Keeping pets or plants	12 (13)	2 (4)	10 (22)	6.13 (1)	.01
Social					
Home visits	46 (50)	24 (51)	22 (48)	0.10 (1)	.75
Calls	91 (98)	46 (98)	45 (98)	<0.001 (1)	.99
Videocalls	45 (48)	23 (49)	22 (48)	0.01 (1)	.91
Texting	46 (50)	25 (53)	21 (46)	0.53 (1)	.47

Knowledge of COVID-19 Situation and Health and Social Services

Of the 93 respondents, 58 (64%) reported accessing moderate or too much COVID-19 information, 89 (96.7%) learned about COVID-19 from television, and 56 (62%) described their understanding of the information as extreme. Moreover, 39/93 (38%) respondents had contacted health and social services, while 29 (31%) requested information regarding these services during the telephone call (Table 4).

Table 4. Differences between the intervention and control groups regarding information and resources pertaining to the COVID-19 pandemic.

Characteristic	Total (N=93), n (%)	Intervention (n=47), n (%)	Control (n=46), n (%)	Chi-square (*df*)	*P* value
Amount of COVID-19[a] information accessed					
None	0 (0)	0 (0)	0 (0)	2.55 (3)	.47
Too little	9 (10)	3 (7)	6 (13)		
Moderate	29 (32)	15 (33)	14 (30)		
Too much	29 (32)	17 (38)	12 (26)		
Extreme	24 (26)	10 (22)	14 (30)		
COVID-19 information source					
Family and friends	47 (51)	27 (59)	20 (43)	2.13 (1)	.14
Television	89 (97)	45 (98)	44 (96)	0.34 (1)	.56
Newspaper	5 (5)	3 (7)	2 (4)	0.21 (1)	.65
Digital media	11 (12)	5 (11)	6 (13)	0.10 (1)	.75
Radio	11 (12)	6 (13)	5 (11)	0.10 (1)	.75
Understanding of COVID-19 information					
None	0 (0)	0 (0)	0 (0)	2.85	.41
Too little	7 (8)	2 (4)	5 (11)		
Moderate	13 (14)	8 (18)	5 (11)		
Too much	15 (17)	9 (20)	6 (13)		
Extreme	56 (62)	26 (58)	30 (65)		
Resources contacted					
None	54 (61)	30 (68)	24 (55)	3.17 (3)	.37
Health services	32 (36)	14 (32)	18 (41)		
COVID-19 services	1 (1)	0 (0)	1 (2)		
Emergency services	0 (0)	0 (0)	0 (0)		
Social services nongovernmental organization	1 (1)	0 (0)	1 (2)		
Resources used					
Health services number	12 (13)	3 (6)	9 (20)	3.59 (1)	.058
COVID-19 services number	1 (1)	0 (0)	1 (2)	-	-
Social services nongovernmental organization number	3 (3)	1 (2)	2 (4)	0.62 (1)	.49
TV-AssistDem Health Education	13 (14)	13 (14)	N/A[b]	N/A	N/A

[a]COVID-19: coronavirus disease.

[b]Not applicable.

Differences Between Living Alone and With Others

In comparison with the participants living with others (69/93, 74%), the participants living alone (24/93, 26%) reported less well-being (35% vs 71%; χ^2=9.61; *P*=.002), more anxiety (59% vs 41%; χ^2=3.90; *P*=.048) and more sleeping problems (48% vs 19%; χ^2=4.71; *P*=.03). They more frequently reported being sad (44% vs 25%; χ^2=2.953; *P*=.09) and bored (26% vs 10%; χ^2=3.613; *P*=.057); however, these last results were only marginally significant.

Differences Between the Intervention and Control Groups

There were no significant differences between the intervention and control groups in any sociodemographic variables, health status variables, or other variables associated with COVID-19 (Tables 1-4). Similarly, there were no differences regarding health management, mental health, well-being, or sleeping problems. Respondents with TV-AssistDem performed more memory exercises than control participants (24/93, 52% vs 8/93, 17%; *P*<.001).

Discussion

Principal Results

Our findings show that at the time of assessment, the physical and mental health and well-being of our study participants with mild cognitive impairment or mild dementia was overall optimal, although living alone was found to be a risk factor for greater psychological negative impact and sleeping problems. Television-based health and social support were provided in the intervention group, and telephone-based support was provided when requested. Television sets stood out as the preferred technological devices to access COVID-19 information, watch television as a recreational activity, and perform memory exercises as an intellectual activity.

Our sample presented characteristics which have been described as preconfinement predictors for greater health and well-being impact: female sex, old age, and mild cognitive impairment or mild dementia [4,5]. However, the change in living arrangements suggests that some households decided to rearrange their support network for the duration of the confinement. Having a support group at home has been described as helpful during disease outbreaks [4]. Our findings are encouraging in that they demonstrate that guaranteeing basic supplies, performing meaningful activities, and ensuring understanding of the situation by providing information from available resources improve the experience of confinement, as described by Brooks et al [4].

Overall, the respondents experienced optimal health status at the time of data collection, which can be explained by the reduced risk of exposure due to decreased daily life outings; they also reported having adequate supplies, which has been reported to mitigate the consequences of quarantine [4]. Our findings regarding mental health and well-being 2 weeks into confinement coincide with those described in the literature for quarantines under 10 days. Our respondents expressed fear, frustration, and boredom, which are frequently expressed negative feelings during confinement [4].

Participation in meaningful activities goes beyond pleasure or entertainment in people with mild cognitive impairment or mild dementia and has shown benefits for cognition and functioning improving independence in instrumental activities of daily living [22]. Physical activity plays a role in enhancing and maintaining cognition [23]; while the outdoor daily life and physical activities of our respondents were restricted, the vast majority engaged daily in physical activities. Leisure activities involving intellectually high cognitive effort or social interaction have been associated with better cognition [24]. Considering that nonattendance of memory workshops and day care services and social isolation may worsen the cognition and functioning of this population, participating in memory games analogically or technologically using smartphones, tablets, computers, or television-based devices may slow the negative consequences of confinement on cognition. The fact that respondents with TV-AssistDem performed significantly more memory exercises than control respondents suggests the potential of television-based activity for cognitive stimulation. Recreational activities have also demonstrated benefits in dealing with challenging situations [24]; however, careful attention must be given to watching television, which is associated with a dose-response cognitive decline [25].

Evidence suggests that access to devices as smartphones, tablets, computers, and television-based devices, which facilitate connectedness and communication, may reduce feelings of isolation [4]; the sample in this study has access to these devices. Since social isolation and loneliness have been associated with poorer cognition [26], special attention must be drawn to people with mild cognitive impairment or mild dementia regarding social activities during confinement. Furthermore, setting up a specific telephone support line has been described as effective in terms of providing health and social support [4,27]. In addition, TV-AssistDem offers and will continue to offer videocall services during confinement to participants in the intervention group.

Inadequate information is generally a potential stressor when facing novel and unknown situations [4]. While poor information may prevent people from comprehending the severity of the situation and complying with the measures, being extremely informed may significantly impact their perception of the situation and cause extreme worry. People with mild cognitive impairment or mild dementia face additional risk, as their cognition and understanding of the information may be compromised. Considering our population characteristics and that their access to COVID-19 information was reported to be moderate to too much, their extreme understanding is exceptional. While television sets stand out as the main source of information, family and friends are the second most common source, which may explain the participants' remarkable understanding of the situation.

To our knowledge, this is the first study to explore the impact of COVID-19 confinement on the physical and mental health and well-being of community-dwelling older adults with mild cognitive impairment or mild dementia and the use of telehealth home support during COVID-19 confinement; no systematic reviews or clinical trials have been registered regarding this study population and technology to date.

Limitations

Although telephone call interviewing was the safest means to communicate with people with mild cognitive impairment or mild dementia during the COVID-19 pandemic, there were several drawbacks. The amount of information gathered and provided in a single telephone call is limited, and researchers were required to balance the time spent on each call. Furthermore, overloading people with mild cognitive impairment or mild dementia with a long interview is not advisable, as it may feel tedious and time-consuming; also, excessive information should not be provided, as they may not be able to understand or remember it all in one telephone call. To ensure telephone call standardization, researchers followed an exhaustive structured protocol.

The impacts on physical and mental health and well-being were assessed 2 weeks into confinement. Studies show that aggravation of physical chronic conditions and poorer mental health and well-being, specifically significantly higher

posttraumatic stress symptoms, appear when a quarantine lasts longer than 10 days. As the duration of the Spanish confinement is not yet clear, the long-term impact and consequences will need to be assessed over time [4]. At the time of submission of this publication, the government had announced an extension of the period of confinement; therefore, new evaluations will be carried out to determine the effects of the prolonged duration of the confinement.

Conclusions

TV-AssistDem, our television-based assistive integrated technology, has been demonstrated to go beyond its initial objective of telehealth home support, promoting active aging of the elderly in their own homes and reducing caregiver burden. It has emerged as a promising cognitive stimulation and telehealth tool to deliver health care and facilitate remote caregiver support during exceptional circumstances, such as the current COVID-19 outbreak.

Our findings suggest that living alone is a risk factor for mental health and well-being and sleep during the coronavirus disease confinement in people with mild cognitive impairment or mild dementia. Recommended measures to address the negative experience of confinement include keeping well informed about the situation and accessing health and social services, having a reliable support network that decreases risk of exposure to COVID-19 and guarantees food and medical supplies, establishing a daily routine with maintained sleeping habits and leisure activities, staying physically and mentally active with cognitive stimulation exercises, and ensuring social connectedness using technology. Our findings show the potential of television sets for informative, recreational, and intellectual purposes in this population.

The findings of this study are valuable and meaningful and contribute to the growing evidence of COVID-19 research across populations, with special attention to people with mild cognitive impairment or mild dementia and technologies. Research regarding the assessment of needs of people with mild cognitive impairment or mild dementia and their caregivers during the COVID-19 pandemic, as well as technology-based support interventions, are urgently globally needed [6,28]. Our approach to explore the uniqueness of this experience in vulnerable populations is appropriate and replicable. Telephone-based interventions during the COVID-19 pandemic to survey and raise health awareness in older adults have proven to be feasible when face-to-face measures are not possible during outbreaks [27].

Government preparedness and responses to situations of the magnitude of COVID-19 determine related outcomes and consequences that go beyond the disease itself and have political, economic, and social impacts. In the short term, this study aims to prepare countries that have yet to face similar governmental restrictions to plan accordingly to guarantee and protect the physical and mental health and well-being of their community-dwelling older adults with mild cognitive impairment or mild dementia. In the long term, this study will contribute to the preparedness for another possible future outbreak. The consequences of this outbreak may affect not only people in confinement and their caregivers but also the health care system, which will need to provide care to address these consequences. The need for remote approaches to outreach and screen people at risk of social isolation may stimulate enhanced implementation of telehealth information and communications technology, such as TV-AssistDem, in the care and support of vulnerable populations who suffer the effects of the COVID-19 pandemic.

Acknowledgments

The authors acknowledge the TV-AssistDem consortium, SmartHealth TV Solution, for the project technology and COVID-19 adaptation and Stimulus for providing the cognitive stimulation software. This project is nested in the TV-AssistDem project, which was selected for funding under the 2016 Call for Proposals focused on “Living well with dementia: The contribution of information and communications technology to integrated solutions for enabling the well-being of people living with dementia and their communities” by the European Active and Assisted Living Programme: TV-AssistDem (AAL-2016-024). This study was additionally funded by Instituto Carlos III through project AC16/00080.

Authors' Contributions

The authorship of this manuscript follows the International Committee of Medical Journal Editors standards. JMGC and EDP made substantial contributions to the conception and design of the work; JMGC and EDP acquired the data; JMGC and JGP analyzed and interpreted the data; JMGC and JGP drafted the work, and JMGC, EDP, JGP, ACV, and FM revised it critically for important intellectual content; and JMGC, EDP, JGP, ACV and FM gave final approval of the version to be published.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey questions.
[DOCX File , 21 KB - jmir_v22i5e19434_app1.docx]

References

1. Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, et al. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). Int J Surg 2020 Apr;76:71-76 [FREE Full text] [doi: 10.1016/j.ijsu.2020.02.034] [Medline: 32112977]
2. Informe sobre la situación de COVID-19 en personal sanitario en España. 2020 Apr 05. Government of Spain URL: https://www.lamoncloa.gob.es/serviciosdeprensa/notasprensa/sanidad14/Documents/2020/060520-Informe%20sobre%20la%20situaci%C3%B3n%20de%20COVID-19%20en%20personal%20sanitario.pdf [accessed 2020-05-14]
3. Legido-Quigley H, Mateos-García JT, Campos VR, Gea-Sánchez M, Muntaner C, McKee M. The resilience of the Spanish health system against the COVID-19 pandemic. Lancet Public Health 2020 May;5(5):e251-e252. [doi: 10.1016/s2468-2667(20)30060-8]
4. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. Lancet 2020 Mar;395(10227):912-920. [doi: 10.1016/s0140-6736(20)30460-8]
5. World Health Organization. 2020 Mar 18. Mental health and psychosocial considerations during the COVID-19 outbreak URL: https://www.who.int/publications-detail/WHO-2019-nCoV-MentalHealth-2020.1 [accessed 2020-05-14]
6. Wang H, Li T, Barbarino P, Gauthier S, Brodaty H, Molinuevo JL, et al. Dementia care during COVID-19. Lancet 2020 Apr;395(10231):1190-1191. [doi: 10.1016/s0140-6736(20)30755-8]
7. Hawton A, Green C, Dickens AP, Richards SH, Taylor RS, Edwards R, et al. The impact of social isolation on the health status and health-related quality of life of older people. Qual Life Res 2011 Feb;20(1):57-67. [doi: 10.1007/s11136-010-9717-2] [Medline: 20658322]
8. Schrempft S, Jackowska M, Hamer M, Steptoe A. Associations between social isolation, loneliness, and objective physical activity in older men and women. BMC Public Health 2019 Jan 16;19(1):74 [FREE Full text] [doi: 10.1186/s12889-019-6424-y] [Medline: 30651092]
9. Kelly ME, Duff H, Kelly S, McHugh Power JE, Brennan S, Lawlor BA, et al. The impact of social activities, social networks, social support and social relationships on the cognitive functioning of healthy older adults: a systematic review. Syst Rev 2017 Dec 19;6(1):259 [FREE Full text] [doi: 10.1186/s13643-017-0632-2] [Medline: 29258596]
10. Kuiper JS, Zuidersma M, Oude Voshaar RC, Zuidema SU, van den Heuvel ER, Stolk RP, et al. Social relationships and risk of dementia: A systematic review and meta-analysis of longitudinal cohort studies. Ageing Res Rev 2015 Jul;22:39-57. [doi: 10.1016/j.arr.2015.04.006] [Medline: 25956016]
11. Braak H, Del Tredici K. Alzheimer's disease: pathogenesis and prevention. Alzheimers Dement 2012 May;8(3):227-233. [doi: 10.1016/j.jalz.2012.01.011] [Medline: 22465174]
12. Alzheimer Europe. Living with dementia: COVID-19 URL: https://www.alzheimer-europe.org/Living-with-dementia/COVID-19 [accessed 2020-05-14]
13. Cotterell N, Buffel T, Phillipson C. Preventing social isolation in older people. Maturitas 2018 Jul;113:80-84. [doi: 10.1016/j.maturitas.2018.04.014] [Medline: 29903652]
14. Smith AC, Thomas E, Snoswell CL, Haydon H, Mehrotra A, Clemensen J, et al. Telehealth for global emergencies: Implications for coronavirus disease 2019 (COVID-19). J Telemed Telecare 2020 Mar 20:1357633X20916567 [FREE Full text] [doi: 10.1177/1357633X20916567] [Medline: 32196391]
15. Government of Spain. Manifiesto: Asistencia COVID-19 URL: https://asistencia.covid19.gob.es/manifiesto [accessed 2020-05-14]
16. Servicio Andaluz de Salud Internet. COVID-19. Información para la ciudadanía URL: https://www.sspa.juntadeandalucia.es/servicioandaluzdesalud/ciudadania/consejos-de-salud/nuevo-coronavirus-informacion-sobre-la-alerta [accessed 2020-05-14]
17. European Commission. Call for proposals AAL 2016 Active and Assisted Living Programme Challenge-Led Call for Proposals AAL 2016 Living well with dementia: Providing integrated solutions based on ICT to support the wellbeing of people living with dementia and their communities URL: http://ec.europa.eu/programmes/horizon2020/en/h2020-section/health-demographic-change-and-wellbeing [accessed 2020-05-14]
18. Goodman-Casanova J, Guzmán-Parra J, Guerrero G, Vera E, Barnestein-Fonseca P, Cortellessa G, et al. TV-based assistive integrated service to support European adults living with mild dementia or mild cognitive impairment (TV-AssistDem): study protocol for a multicentre randomized controlled trial. BMC Geriatr 2019 Sep 06;19(1):247 [FREE Full text] [doi: 10.1186/s12877-019-1267-z] [Medline: 31492113]
19. Khosravi P, Ghapanchi AH. Investigating the effectiveness of technologies applied to assist seniors: A systematic literature review. Int J Med Inform 2016 Jan;85(1):17-26. [doi: 10.1016/j.ijmedinf.2015.05.014] [Medline: 26216463]
20. Calton B, Abedini N, Fratkin M. Telemedicine in the Time of Coronavirus. J Pain Symptom Manage 2020 Mar 31:1-3. [doi: 10.1016/j.jpainsymman.2020.03.019] [Medline: 32240756]
21. Gordon M. Assess Notes: Nursing Assessment and Diagnostic Reasoning. Philadelphia, PA: FA Davis Co; 2008.
22. Nyman SR, Szymczynska P. Meaningful activities for improving the wellbeing of people with dementia: beyond mere pleasure to meeting fundamental psychological needs. Perspect Public Health 2016 Mar;136(2):99-107. [doi: 10.1177/1757913915626193] [Medline: 26933079]

23. Etnier J, Drollette E, Slutsky A. Physical activity and cognition: A narrative review of the evidence for older adults. Psychol Sport Exerc 2019 May;42:156-166. [doi: 10.1016/j.psychsport.2018.12.006]
24. Singh-Manoux A, Richards M, Marmot M. Leisure activities and cognitive function in middle age: evidence from the Whitehall II study. J Epidemiol Community Health 2003 Nov;57(11):907-913 [FREE Full text] [doi: 10.1136/jech.57.11.907] [Medline: 14600119]
25. Fancourt D, Steptoe A. Television viewing and cognitive decline in older age: findings from the English Longitudinal Study of Ageing. Sci Rep 2019 Feb 28;9(1):2851 [FREE Full text] [doi: 10.1038/s41598-019-39354-4] [Medline: 30820029]
26. O'Luanaigh C, O'Connell H, Chin A, Hamilton F, Coen R, Walsh C, et al. Loneliness and cognition in older people: the Dublin Healthy Ageing study. Aging Ment Health 2012;16(3):347-352. [doi: 10.1080/13607863.2011.628977] [Medline: 22129350]
27. Chan SS, So WK, Wong DC, Lee AC, Tiwari A. Improving older adults' knowledge and practice of preventive measures through a telephone health education during the SARS epidemic in Hong Kong: a pilot study. Int J Nurs Stud 2007 Sep;44(7):1120-1127 [FREE Full text] [doi: 10.1016/j.ijnurstu.2006.04.019] [Medline: 16857203]
28. Comas-Herrera A, Lorenz-Dant K, Ferri C, Govia I, Sani TP, Jacobs R, et al. LTC Responses to COVID-19. 2020 Apr 10. Supporting people living with dementia and their carers in low-and middle-income countries during COVID-19 URL: https://ltccovid.org/wp-content/uploads/2020/04/COVID-and-people-with-dementia-in-low-and-middle-income-countries.pdf [accessed 2020-05-14]

Abbreviations

AAL: active and assisted living
COVID-19: coronavirus disease
ICU: intensive care unit
TV-AssistDem: TeleVision-based ASSistive Integrated Service to supporT European adults living with mild DEMentia or mild cognitive impairment
WHO: World Health Organization

Edited by G Eysenbach; submitted 17.04.20; peer-reviewed by C Snoswell, E Burner, I Shubina; comments to author 04.05.20; revised version received 11.05.20; accepted 12.05.20; published 22.05.20.

Please cite as:
Goodman-Casanova JM, Dura-Perez E, Guzman-Parra J, Cuesta-Vargas A, Mayoral-Cleries F
Telehealth Home Support During COVID-19 Confinement for Community-Dwelling Older Adults With Mild Cognitive Impairment or Mild Dementia: Survey Study
J Med Internet Res 2020;22(5):e19434
URL: http://www.jmir.org/2020/5/e19434/
doi:10.2196/19434
PMID:32401215

Original Paper

Use of a Real-Time Locating System for Contact Tracing of Health Care Workers During the COVID-19 Pandemic at an Infectious Disease Center in Singapore: Validation Study

Hanley J Ho[1], MPH; Zoe Xiaozhu Zhang[1], PhD; Zhilian Huang[1], PhD; Aung Hein Aung[1], MPH; Wei-Yen Lim[1], PhD; Angela Chow[1,2], PhD

[1]Department of Clinical Epidemiology, Office of Clinical Epidemiology, Analytics, and Knowledge (OCEAN), Tan Tock Seng Hospital, Singapore
[2]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore

Corresponding Author:
Hanley J Ho, MPH
Department of Clinical Epidemiology
Office of Clinical Epidemiology, Analytics, and Knowledge (OCEAN)
Tan Tock Seng Hospital
11 Jalan Tan Tock Seng
308433
Singapore
Phone: 65 6511 5054
Email: hanley_ho@ttsh.com.sg

Abstract

Background: In early 2020, coronavirus disease (COVID-19) emerged and spread by community and nosocomial transmission. Effective contact tracing of potentially exposed health care workers is crucial for the prevention and control of infectious disease outbreaks in the health care setting.

Objective: This study aimed to evaluate the comparative effectiveness of contact tracing during the COVID-19 pandemic through the real-time locating system (RTLS) and review of the electronic medical record (EMR) at the designated hospital for COVID-19 response in Singapore.

Methods: Over a 2-day study period, all admitted patients with COVID-19, their ward locations, and the health care workers rostered to each ward were identified to determine the total number of potential contacts between patients with COVID-19 and health care workers. The numbers of staff-patient contacts determined by EMR reviews, RTLS-based contact tracing, and a combination of both methods were evaluated. The use of EMR-based and RTLS-based contact tracing methods was further validated by comparing their sensitivity and specificity against self-reported staff-patient contacts by health care workers.

Results: Of 796 potential staff-patient contacts (between 17 patients and 162 staff members), 104 (13.1%) were identified by both the RTLS and EMR, 54 (6.8%) by the RTLS alone, and 99 (12.4%) by the EMR alone; 539 (67.7%) were not identified through either method. Compared to self-reported contacts, EMR reviews had a sensitivity of 47.2% and a specificity of 77.9%, while the RTLS had a sensitivity of 72.2% and a specificity of 87.7%. The highest sensitivity was obtained by including all contacts identified by either the RTLS or the EMR (sensitivity 77.8%, specificity 73.4%).

Conclusions: RTLS-based contact tracing showed higher sensitivity and specificity than EMR review. Integration of both methods provided the best performance for rapid contact tracing, although technical adjustments to the RTLS and increasing user compliance with wearing of RTLS tags remain necessary.

(J Med Internet Res 2020;22(5):e19437) doi:10.2196/19437

KEYWORDS

infectious disease; real-time locating systems; electronic medical records; COVID-19; contact tracing; health care workers; RFID

Introduction

In early 2020, coronavirus disease (COVID-19) emerged in Wuhan, China; the disease spread by community and nosocomial transmission, infecting up to 3019 health care workers by February 12, 2020 [1].

Contact tracing is used to identify individuals potentially exposed to infectious diseases; it is crucial for the prevention and control of infectious disease outbreaks [2,3]. Health care workers face high risks of contracting infectious diseases due to the large amounts of time they spend interacting with patients and their coworkers [4,5]. Health care workers experience contact with 14-18 persons in a typical work shift in a general ward, with nurses having the longest duration of physical contact with patients [6]. As contacts in health care settings tend to be close, any exposure to an infectious patient requires immediate contact tracing and contact management. Failure to identify potentially exposed contacts puts other patients and health care workers at greater risk of infection [7].

Conventional contact tracing methods are limited by their cost and reliability. Continuous direct observation has been considered to be the gold standard method to accurately quantify contact time (including activities and locations of interest); however, due to the intensive human resource requirements of this method, it is cost-ineffective and impractical for large-scale projects [8]. Self-reporting methods, such as activity diaries and interviews, have been used as alternatives to direct observation due to the lower intensity of their human resource demands [6,8]; however, these methods are also time-consuming and subject to reporting biases that compromise the accuracy of the data collected [9,10]. Another commonly used method for contact tracing is data extraction from administrative and clinical databases such as electronic medical records (EMRs) [8,11]. Although this method overcomes some of the problems related to cost and reliability, it is also time-consuming and limits the capture of patient contact episodes to care providers who can input data in the EMR [8,12].

Technological advances such as real-time locating systems (RTLSs) have shown promise in overcoming the barriers to conventional contact tracing methods [13,14]. In recent years, health care institutions have been increasingly exploring the use of RTLSs to establish contacts within health care premises [12,13,15]. One type of RTLS technology is radio-frequency identification (RFID) tracking. RFID tracking requires the user to wear an RFID tag, which continuously sends wireless signals to sensors (RFID readers) installed at various locations in the hospital [16]. RFID technology can provide an accurate gauge of health care workers' movements and their interactions with patients and coworkers [15]. This capability is useful for hospital contact tracing during an infectious disease outbreak.

While many studies have been conducted on community-based contact tracing measures using digital technology, fewer articles have focused on technology for hospital-based contact tracing for health care workers. During the COVID-19 pandemic, we evaluated the comparative effectiveness of contact tracing through RTLSs and the conventional method of identifying contacts from EMRs and validated these methods against self-reporting of contacts by health care workers.

Methods

Setting

This study was performed during the COVID-19 pandemic in Singapore at the National Centre for Infectious Diseases (NCID), the designated hospital for COVID-19 response. Singapore identified its first COVID-19 case on January 23, 2020; by February 19, 2020, a total of 84 cases were reported nationally [17]. Of these, the NCID managed 65 cases. The study was conducted over 2 days from midnight on February 25, 2020 to 11:59 PM on February 26, 2020. The NCID is a 330-bed purpose-built facility for the management of emerging infectious diseases. The NCID building was outfitted with RTLS technology, and location trackers were installed in all ward areas. All staff working within the building were provided with RTLS tags, which they routinely carried during work. The tags also served as access cards to the NCID building and inpatient wards.

Real-Time Locating System

Staff wearing RTLS tags could be located within the wards and inpatient rooms of the NCID. All inpatient rooms were fitted with RTLS location exciters and wireless access points. Whenever a tag passed a location exciter in an inpatient room, the tag would receive a low-frequency signal and transmit a radio frequency signal to the access point, where the location triangulation technology would decipher the signals and determine the exact location of the RTLS tag.

Electronic Medical Record

The NCID also uses an EMR system to capture clinical encounters of all inpatients. Staff providing clinical care were issued with personal accounts to make entries into patients' clinical notes regarding their clinical assessments, medication orders, laboratory and radiological tests, and charting of vital signs.

Participants

Participants included medical and nursing staff who could have had contact with patients with COVID-19 isolated in the general wards of the NCID during the 2-day study period. Inpatients with a laboratory-confirmed diagnosis of COVID-19 infection were identified through the hospital's laboratory information system, and on-duty medical and nursing staff were identified from duty rosters.

We excluded staff members who did not have access to the hospital EMR system, did not have a working RTLS tag associated with their identity, or did not have a sufficiently charged RTLS tag that could be reliably detected by the location trackers. We identified these staff members by checking all staff that were on the roster for any movement records captured by the RTLS in the week preceding the 2 days of the study, and we excluded those whose identity could not be found in the RTLS (ie, they were not issued a tag) or who did not register any movement record (ie, they were not using their tag or their tag had not been charged).

Study Design

The study was divided into two parts. In the first part of the study, we compared two methods of contact tracing for medical and nursing staff who had come into contact with a confirmed patient with COVID-19 over the study period: the conventional method of reviewing the EMR and the extraction of staff records from the RTLS. For the EMR reviews, we extracted the names of all physicians and nurses who had come into contact with patients based on entries in the EMR. Records included medical and nursing notes as well as documentation of patients' vital signs. For extraction of staff records from the RTLS, we used location-based tracking to identify all individuals picked up by the appropriate location tracker in a specific airborne infection isolation room to which each patient with COVID-19 was admitted. We used a highly sensitive cutoff of at least 1 second (ie, as long as a staff member's RTLS tag was picked up by the location tracker in a specific room, we considered that the staff member had been in contact with the patient admitted to that room).

For each patient day (a 24-hour period from midnight to 11:59 PM for each specific patient with COVID-19), we counted the total number of unique staff-patient contacts identified using EMR-based and RTLS-based contact tracing methods. For staff with multiple contacts with the same patient in the same day, we included only the first contact episode. We then classified the episodes into contacts identified through both methods and contacts identified through either one of the two methods. We constructed 2×2 tables for each patient day and aggregated these tables. We reported the total number of contacts identified by the two methods, the proportions identified by both methods, and the proportions identified by either one of the two methods.

In the second part of our study, we attempted to validate both methods against self-reporting by physicians and nurses. For this part of the study, we restricted our analysis only to patients with COVID-19 who had been admitted to two wards in the NCID on February 25, 2020. We identified all possible contacts of a patient by examining the rostered medical and nursing staff for a given ward. We then separately used both the EMR review and RTLS staff record extraction methods as detailed above to identify all medical and nursing staff members who had come into contact with each patient with COVID-19. For the self-reporting method, we contacted all physicians and nursing staff members rostered to those wards by telephone and asked them if they had physically entered the airborne infection isolation room (not just the anteroom) of each patient with COVID-19. This question was asked on the day after the day of interest to reduce recall bias. Our hospital protocol requires all staff managing patients with COVID-19 to don full personal protection equipment (including a fit-tested N95 mask, gown, gloves, and goggles/face shield) during their encounters with patients. This requirement is strictly enforced in the wards by both the ward nursing manager and senior physicians. As such, these staff members were not in danger of being quarantined or of being reprimanded for not following protocol. We therefore expected these self-reports to be truthful.

For each patient day, we constructed two 2×2 tables: one comparing the performance of the EMR review method against self-reported contacts (ie, its comparative ability to correctly identify staff-patient contacts and non–staff-patient contacts), and the other comparing the RTLS method against self-reported contacts. Tables for all patient days were then aggregated. We used the aggregated tables to calculate the sensitivity (proportion of correctly identified staff-patient contacts) and specificity (proportion of correctly identified non–staff-patient contacts) of both methods separately. In addition, we created three logistic regression models in which the dependent variable was the self-reported contacts: in Model 1, the only independent variable was EMR detection, in Model 2, the only independent variable was RTLS detection, and in Model 3, both variables were included as independent variables. We then used the likelihood ratio test to compare the goodness of fit of these models.

We further considered the sensitivity and specificity of using both methods concurrently, using either an "RTLS or EMR" approach or an "RTLS and EMR" approach, to detect staff-patient contacts.

Results

RTLS Versus EMR-Based Contact Tracing

Our study included 17 inpatients with COVID-19 warded at the NCID on February 25 and 26, 2020, housed in single airborne infection isolation rooms across six isolation wards. From the ward duty rosters, a total of 212 staff members (30 medical, 14.2%, and 182 nursing, 85.8%) were rostered for duty in these six wards over the study period. We excluded 50/212 staff members (23.6%; 5 medical, 10.0%, and 45 nursing, 90.0%) due to tag-related issues, and the remaining 162 staff members (25 medical, 15.4%, and 137 nursing, 84.6%) were included in our study. In total, based on the ward location of each patient and the number of staff rostered to each ward over the 2-day study period, 796 potential staff-patient contacts were identified between these 17 inpatients and 162 staff members.

Table 1 compares the number of contacts identified by the RTLS with those identified by the EMR for 34 patient days. Of 796 potential staff-patient contacts, 104 (13.1%) were identified by both RTLS and EMR, 54 (6.8%) by RTLS alone, and 99 (12.4%) by EMR alone; 539 (67.7%) were not identified through either method. Among the total of 257 staff-patient contacts identified, the RTLS identified 158 (61.5%), while the EMR identified 203 (79.0%). However, it is not possible to determine whether the contacts were identified accurately.

Table 1. Summary of possible contacts identified by the RTLS and EMR over 34 patient days. Percentages are calculated according to row values.

EMR detection status	Detected by RTLS[a]	Not detected by RTLS	Total
Detected by EMR[b]	104 (13.1%)	99 (12.4%)	203
Not detected by EMR	54 (6.8%)	539 (67.7%)	593
Total	158	638	796

[a]RTLS: real-time locating system.

[b]EMR: electronic medical record.

Validation of RTLS-Based and EMR-Based Contact Tracing

For our validation study, we evaluated staff-patient contacts for 10 confirmed patients with COVID-19 in two wards (5 in Ward A and 5 in Ward B). During the 1-day validation study period, 36 staff members (6 medical, 17%, and 30 nursing, 83%) were rostered to Ward A and 30 staff members (6 medical, 20%, and 24 nursing, 80%) were rostered to Ward B. Of these, 8 staff members (1 medical, 13%, and 7 nursing, 87%) from Ward A and 2 staff members (both nursing) from Ward B were excluded due to staff tag–related issues. Therefore, we included 28 staff members from Ward A and 28 staff members from Ward B in our study. This gave a total of 280 potential staff-patient contacts.

In general, EMR review produced more staff records than the RTLS (P<.001). The overall observed agreement between the RTLS and EMR was 80.8%.

Table 2 compares the performance of RTLS staff records with self-reported contacts. Of 280 potential staff-patient contacts, 36 (12.9%) were self-reported as having occurred. Of these, 26 contacts were traced by RTLS, giving a sensitivity of 72.2%. Of 244 self-reported non–staff-patient contacts, 214 were accurately identified by RTLS, giving a specificity of 87.7%. The positive predictive value was 46.4%, while the negative predictive value was 95.5%.

Table 3 compares the performance of EMR review against self-reporting by staff. Of 36 self-reported staff-patient contacts, 17 were identified by EMR review, giving a sensitivity of 47.2%. Of 244 staff-patient contacts not reported by staff, 190 were identified by EMR review, giving a specificity of 77.9%. The positive predictive value was 23.9%, while the negative predictive value was 90.9%.

Table 2. Comparison of the performance of RTLS-based contact tracing with self-reported contacts with patients with COVID-19. Percentages are calculated according to row values.

Self-reported status	Detected by RTLS[a]	Not detected by RTLS	Total
Contacts by self-report	26 (72.2%)	10 (27.8%)	36
Noncontacts by self-report	30 (12.3%)	214 (87.7%)	244
Total	56	224	280

[a]RTLS: real-time locating system.

Table 3. Comparison of the performance of EMR-based contact tracing against self-reported contacts with patients with COVID-19. Percentages are calculated according to row values.

Self-reported status	Detected by EMR[a]	Not detected by EMR	Total
Contacts by self-report	17 (47.2%)	19 (52.8%)	36
Noncontacts by self-report	54 (22.1%)	190 (77.9%)	244
Total	71	209	280

[a]EMR: electronic medical record.

Among the logistic regression models, Model 3 performed significantly better than Model 1 (P<.001) but not differently from Model 2 (P=0.84) (Table 4). These results suggest that the RTLS was better than the EMR at identifying self-reported contacts. This is consistent with our results showing the higher sensitivity and specificity of the RTLS than of the EMR compared with self-reported contacts.

Table 5 shows the overall performance of combining RTLS-based and EMR-based contact tracing. If we used an "or" strategy and considered staff-patient contacts to have truly occurred if they were traced by either RTLS or EMR review, the sensitivity increased to 28/36 (77.8%); however, this came at a cost of lower specificity (179/244, 73.4%). On the other hand, if we used an "and" strategy and considered staff-patient contacts to have truly occurred if they were traced by both RTLS and EMR, the sensitivity decreased to 15/36 (41.7%) while the specificity increased to 225/244 (92.2%).

Table 4. Coefficients and comparisons of three logistic regression models for the EMR method, the RTLS method, and both methods.

Comparison	Model 1[a]	Model 2[b]	Model 3[c]
EMR[d] detection, coefficient (95% CI)	1.15 (0.43 to 1.87)	N/A[e]	–0.10 (–1.01 to 0.82)
RTLS[f] detection, coefficient (95% CI)	N/A	2.92 (2.40 to 3.74)	2.96 (2.04 to 3.88)
P value of nested model comparison[g] with Model 3	.001	.84	N/A

[a]Model 1: EMR detection is the only independent variable.

[b]Model 2: RTLS detection is the only independent variable.

[c]Model 3: both EMR detection and RTLS detection are independent variables.

[d]EMR: electronic medical record.

[e]Not applicable.

[f]RTLS: real-time locating system.

[g]Compared by likelihood ratio test.

Table 5. Comparison of two different approaches to combining EMR and RTLS data versus self-reported contacts with patients with COVID-19. Percentages are calculated according to row values.

Self-reported status	Contact tracing using an "or" strategy		Contact tracing using an "and" strategy		Total
	Detected by EMR[a] or RTLS[b]	Not detected by EMR or RTLS	Detected by EMR and RTLS	Not detected by EMR and RTLS	
Contacts by self-report	28 (77.8%)	8 (22.2%)	15 (41.7%)	21 (58.3%)	36
Noncontacts by self-report	65 (26.6%)	179 (73.4%)	19 (7.8%)	225 (92.2%)	244
Total	93	187	34	246	280

[a]EMR: electronic medical record.

[b]RTLS: real-time locating system.

Discussion

Principal Findings

Timely and accurate hospital contact tracing is vital to preserve staff and patient safety and to prevent nosocomial transmission of infectious diseases. Our study demonstrated that RTLS-based contact tracing has higher sensitivity and specificity than EMR-based contact tracing compared to self-reporting by staff.

The performance of the EMR reviews was surprisingly poor, with low sensitivity of 47% and moderate specificity of <78%. This is likely related to clinical operational processes, where the staff member performing clinical documentation for the patient may not be the same staff member who entered the patient's room. This is especially likely in the context of a busy airborne infection isolation ward, where staff may split the workload of performing practical procedures (which would require the donning and doffing of PPE and time interacting with patients to perform clinical procedures) and administrative tasks (including clinical documentation). A lack of clarity was likely present when documenting which specific staff members had performed certain tasks. Moreover, staff-patient contacts that are not required to be noted in routine reports or that did not significantly affect clinical care of the patient may not have been documented.

Consequently, although EMR review detected more staff-patient contacts than the RTLS over our 2-day study period, many of the contacts would have either been false positive or false negative. The inability to accurately contact trace staff may lead to false alarms or false reassurances when evaluating the risk of exposure to an infectious patient.

The RTLS-based contact tracing method performed better than the EMR reviews, with moderate sensitivity of 72% and high specificity of 88%. While we did not specifically measure the time taken for each method, it was estimated that the RTLS data extraction time was approximately 2 to 3 minutes per patient, while the EMR record reviews required approximately 20 minutes per patient. These findings are comparable to previous studies validating the accuracy and ease of RFID technology in quantifying human contact episodes [12,18,19]. Chang et al [19] validated the accuracy of RFID tag readers (>80%) in detecting proximity events in the intensive care unit by comparing data obtained from direct observation; they demonstrated sensitivities ranging from 73.8%-90.9% as well as specificities ranging from 83.8%-98.0% for the technology used, with better performance for invasive events. Lucet et al [18] found no difference in the interaction duration between health care workers and patients with tuberculosis in airborne isolation when comparing records obtained from RFID network sensors, direct observations, and interviews. Hellmich et al [12] showed that the use of RFID technology in the emergency department generated twice as many contacts compared with the conventional method of EMR review during a pertussis outbreak, and each RTLS data query required less than 5 minutes, compared to 30-60 minutes per EMR review.

To identify the number of staff members in contact with an infectious patient as accurately as possible, high test sensitivity is desired. Based on our study, RTLS-based contact tracing would be able to identify most contacts with exposure to a patient with COVID-19. The sensitivity of this method could be increased slightly by integrating EMR-based contact tracing methods (with a slight compromise of specificity). Presently, this may be the most practicable solution to obtain contacts quickly. Further verifications with staff would be required to capture the most accurate list of staff contacts requiring follow-up and intervention.

Aside from staff-patient contacts, there is also potential to use RTLS technology to determine staff-staff contacts by using the RTLS data to analyze which staff members were in the same predemarcated zones at the same time. This has important practical applications in the event that a health care worker is confirmed to have an infectious disease and contact tracing of exposed staff is required.

The results obtained in our study were partially limited by implementation-related challenges of the RTLS system. RTLS electronic tags held by each staff member would need to be detected by the correct location exciter when the staff member moved from one demarcated zone to another. Careful calibration of the technology by the developers was necessary to optimize the contact tracing results.

We also observed that it was necessary to exclude a sizeable proportion of staff from our study due to staff tag–related issues. This is partly because in the setting of the COVID-19 outbreak, some nursing staff were rapidly redeployed to augment the personnel at the NCID; consequently, RTLS tags were not readily available to them. Furthermore, issues related to staff acceptance of RTLS technology, compliance with carrying the tag consistently during routine work, and technical issues such as proper tag association and regular charging of the battery may have been present. Key principles for gaining user acceptance of such technology include clearly conveying the purpose and intended uses of the technology and ensuring that individual electronic tags are practically convenient to use [13]. Although effort was made to ensure that staff are aware of the importance of contact tracing and to integrate door access into a single tag for convenient use, some staff members still did not have fully charged working tags. Further studies are needed to evaluate the knowledge, attitudes, practices, and behaviors of health care workers toward RTLS technology.

While direct observation and self-reporting methods are regarded as having a "higher" standard to accurately determine the duration of contact between staff and patients as well as the nature of the interaction (eg, multiple brief episodes versus a few episodes of prolonged contact) [6,8], such methods are not practically implementable for extended periods of time due to heavy personnel and time requirements (eg, the need for research staff to observe health care workers or for health care workers to complete self-reported diaries for an entire shift of 8-10 hours per day [6]). Other forms of technology, such as closed circuit television monitoring, have also been used to assess risk exposure of movements of health care workers in an outbreak setting [20,21]; however, these technologies are potentially time-intensive and labor-intensive.

Studies validating the accuracy of RFID technology have found no difference in the interaction duration between health care workers and tuberculosis patients when comparing records obtained from RFID network sensors, direct observations, and interviews [18] and in the detection of proximity events in the intensive care unit by comparing data from direct observation [19]. We believe that the RTLS system, when properly implemented, will be a pragmatic means of capturing all staff-patient contact episodes in a similar fashion to facilitate rapid contact management of health care workers.

Strengths

This was a pragmatic study conducted in an outbreak setting; hence, it offers a "real-world" perspective on the usefulness of EMR-based and RTLS-based contact tracing methods. The study team employed a systematic and standardized process to perform the contact tracing using the two methods and to conduct telephone conversations with health care workers to ensure accurate capture of data. Furthermore, the findings of our study add to the limited information on the comparative effectiveness of RFID technology and conventional methods for contact tracing. Most studies on RTLSs are confined to the measurement of human contact duration in health care settings [5,22,23].

Limitations

For the RTLS-based contact tracing, we only used location-based tracking to determine contacts and did not assess the proximity of the contacts. However, we reason that staff members entering the isolation room would likely have needed to be in close contact with the patient (within 2 meters) to perform their clinical duties; hence, this was a reasonable means of determining close contacts with the patient. Regarding our chosen standard of self-reported contacts by health care workers, information bias was possible, as staff were required to recall their movements during a shift that had already been completed. We sought to minimize errors in recall by contacting the participants 1 day after their working shifts and by corroborating their reports with other colleagues on the team. Our study period was short due to resource challenges faced by the study team, whose members were concurrently involved in outbreak management work during the COVID-19 pandemic. However, we believe that our study methodology was robust and that our observations remain valid in our evaluation of novel technology under actual outbreak conditions.

Conclusions

We have demonstrated that RTLS-based contact tracing has higher sensitivity and specificity than EMR-based contact tracing compared with self-reported contacts. Integration of both methods appeared to provide the best performance for rapid contact tracing during the COVID-19 pandemic, with a sensitivity of 78% and a specificity of 73%. Technical adjustments and increasing user compliance are necessary to further improve the effectiveness of the RTLS for contact tracing purposes.

Acknowledgments

No funding was received.

Authors' Contributions

HJH conceived the study, analyzed and interpreted the data, and drafted the manuscript with input from all authors. ZXZ assisted with acquiring and interpreting study data and drafting the manuscript. ZH assisted with interpreting the study data and drafting the manuscript. AAH assisted with analyzing and interpreting the study data and provided input for the manuscript. WL conceived the study, assisted with acquiring, analyzing, and interpreting the study data, and provided input for the manuscript. AC conceived the study, provided overall direction and planning for the study, analyzed and interpreted the data, and critically revised the manuscript. All authors reviewed and approved the final version of the manuscript prior to submission.

Conflicts of Interest

None declared.

References

1. Wang Y, Wang Y, Chen Y, Qin Q. Unique epidemiological and clinical features of the emerging 2019 novel coronavirus pneumonia (COVID-19) implicate special control measures. J Med Virol 2020 Mar 05:568-576. [doi: 10.1002/jmv.25748] [Medline: 32134116]
2. World Health Organization. 2017 May 09. Contact tracing URL: https://www.who.int/features/qa/contact-tracing/en/ [accessed 2020-03-17]
3. Saurabh S, Prateek S. Role of contact tracing in containing the 2014 Ebola outbreak: a review. Afr Health Sci 2017 Mar;17(1):225-236 [FREE Full text] [doi: 10.4314/ahs.v17i1.28] [Medline: 29026397]
4. Kuster SP, Shah PS, Coleman BL, Lam P, Tong A, Wormsbecker A, et al. Incidence of influenza in healthy adults and healthcare workers: a systematic review and meta-analysis. PLoS One 2011;6(10):e26239 [FREE Full text] [doi: 10.1371/journal.pone.0026239] [Medline: 22028840]
5. Lowery-North DW, Hertzberg VS, Elon L, Cotsonis G, Hilton SA, Vaughns CF, et al. Measuring social contacts in the emergency department. PLoS One 2013;8(8):e70854 [FREE Full text] [doi: 10.1371/journal.pone.0070854] [Medline: 23990915]
6. Jiang L, Ng HL, Ho HJ, Leo YS, Prem K, Cook AR, et al. Contacts of healthcare workers, patients and visitors in general wards in Singapore. Epidemiol Infect 2017 Sep 08;145(14):3085-3095. [doi: 10.1017/s0950268817002035]
7. Drumright L, Holmes A. Monitoring major illness in health care workers and hospital staff. Clin Infect Dis 2011 Aug 01;53(3):284-286. [doi: 10.1093/cid/cir384] [Medline: 21765077]
8. Jones TL, Schlegel C. Can real time location system technology (RTLS) provide useful estimates of time use by nursing personnel? Res Nurs Health 2014 Feb;37(1):75-84 [FREE Full text] [doi: 10.1002/nur.21578] [Medline: 24338915]
9. Farrahi K, Emonet R, Cebrian M. Epidemic contact tracing via communication traces. PLoS One 2014;9(5):e95133 [FREE Full text] [doi: 10.1371/journal.pone.0095133] [Medline: 24787614]
10. Foxwell AR, Roberts L, Lokuge K, Kelly PM. Transmission of influenza on international flights, may 2009. Emerg Infect Dis 2011 Jul;17(7):1188-1194 [FREE Full text] [doi: 10.3201/eid1707.101135] [Medline: 21762571]
11. Alamo ST, Wagner GJ, Sunday P, Wanyenze RK, Ouma J, Kamya M, et al. Electronic medical records and same day patient tracing improves clinic efficiency and adherence to appointments in a community based HIV/AIDS care program, in Uganda. AIDS Behav 2012 Feb;16(2):368-374 [FREE Full text] [doi: 10.1007/s10461-011-9996-9] [Medline: 21739285]
12. Hellmich TR, Clements CM, El-Sherif N, Pasupathy KS, Nestler DM, Boggust A, et al. Contact tracing with a real-time location system: A case study of increasing relative effectiveness in an emergency department. Am J Infect Control 2017 Dec 01;45(12):1308-1311 [FREE Full text] [doi: 10.1016/j.ajic.2017.08.014] [Medline: 28967513]
13. Yao W, Chu C, Li Z. The adoption and implementation of RFID technologies in healthcare: a literature review. J Med Syst 2012 Dec;36(6):3507-3525. [doi: 10.1007/s10916-011-9789-8] [Medline: 22009254]
14. Ebrahimzadeh F, Nabovati E, Hasibian MR, Eslami S. Evaluation of the Effects of Radio-Frequency Identification Technology on Patient Tracking in Hospitals: A Systematic Review. J Patient Saf 2017 Dec 15:Online ahead of print. [doi: 10.1097/PTS.0000000000000446] [Medline: 29252967]
15. Bowen ME, Wingrave CA, Klanchar A, Craighead J. Tracking technology: lessons learned in two health care sites. Technol Health Care 2013;21(3):191-197. [doi: 10.3233/THC-130738] [Medline: 23792793]
16. Camacho-Cogollo J, Bonet I, Iadanza E. Chapter 4 - RFID technology in health care. In: Iadanza E, editor. Clinical Engineering Handbook. Cambridge, MA: Academic Press; Dec 04, 2019:A-41.
17. Wong JEL, Leo YS, Tan CC. COVID-19 in Singapore-Current Experience: Critical Global Issues That Require Attention and Action. JAMA 2020 Feb 20:1243-1244. [doi: 10.1001/jama.2020.2467] [Medline: 32077901]

18. Lucet J, Laouenan C, Chelius G, Veziris N, Lepelletier D, Friggeri A, et al. Electronic sensors for assessing interactions between healthcare workers and patients under airborne precautions. PLoS One 2012;7(5):e37893 [FREE Full text] [doi: 10.1371/journal.pone.0037893] [Medline: 22662245]
19. Chang Y, Syed-Abdul S, Tsai C, Li Y. A novel method for inferring RFID tag reader recordings into clinical events. Int J Med Inform 2011 Dec;80(12):872-880. [doi: 10.1016/j.ijmedinf.2011.09.006] [Medline: 22018605]
20. Park Y, Lee C, Kim KM, Kim SW, Lee K, Ahn J, et al. The first case of the 2015 Korean Middle East Respiratory Syndrome outbreak. Epidemiol Health 2015;37:e2015049 [FREE Full text] [doi: 10.4178/epih/e2015049] [Medline: 26725226]
21. Ki HK, Han SK, Son JS, Park SO. Risk of transmission via medical employees and importance of routine infection-prevention policy in a nosocomial outbreak of Middle East respiratory syndrome (MERS): a descriptive analysis from a tertiary care hospital in South Korea. BMC Pulm Med 2019 Oct 30;19(1):190 [FREE Full text] [doi: 10.1186/s12890-019-0940-5] [Medline: 31666061]
22. Isella L, Romano M, Barrat A, Cattuto C, Colizza V, Van den Broeck W, et al. Close encounters in a pediatric ward: measuring face-to-face proximity and mixing patterns with wearable sensors. PLoS One 2011 Feb 28;6(2):e17144 [FREE Full text] [doi: 10.1371/journal.pone.0017144] [Medline: 21386902]
23. Arunachalam S, Asan O, Nestler D, Heaton H, Hellmich T, Wutthisirisart P, et al. Patient-Care Team Contact Patterns Impact Treatment Length of Stay in the Emergency Department. In: Conf Proc IEEE Eng Med Biol Soc. 2019 Jul Presented at: 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 23-27 July 2019; Berlin, Germany p. 345-348. [doi: 10.1109/EMBC.2019.8857803]

Abbreviations

COVID-19: coronavirus disease
EMR: electronic medical record
NCID: National Centre for Infectious Diseases
RFID: radio-frequency identification device
RTLS: real-time locating system

Edited by G Eysenbach; submitted 17.04.20; peer-reviewed by D Parry, C Turcu, J Li; comments to author 13.05.20; revised version received 14.05.20; accepted 14.05.20; published 26.05.20.

Please cite as:
Ho HJ, Zhang ZX, Huang Z, Aung AH, Lim WY, Chow A
Use of a Real-Time Locating System for Contact Tracing of Health Care Workers During the COVID-19 Pandemic at an Infectious Disease Center in Singapore: Validation Study
J Med Internet Res 2020;22(5):e19437
URL: http://www.jmir.org/2020/5/e19437/
doi:10.2196/19437
PMID:32412416

Proposal

A Global Digital Citizen Science Policy to Tackle Pandemics Like COVID-19

Tarun R Katapally[1], PhD

University of Regina, Regina, SK, Canada

Corresponding Author:
Tarun R Katapally, PhD
University of Regina
2155 College Ave
Johnson Shoyama Graduate School of Public Policy
Regina, SK, S4P4V5
Canada
Phone: 1 3065854544
Email: tarun.katapally@uregina.ca

Abstract

The coronavirus disease (COVID-19) pandemic is an extremely complex existential threat that requires cohesive societal effort to address health system inefficiencies. When our society has faced existential crises in the past, we have banded together by using the technology at hand to overcome them. The COVID-19 pandemic is one such threat that requires not only a cohesive effort, but also enormous trust to follow public health guidelines, maintain social distance, and share necessities. However, are democratic societies with civil liberties capable of doing this? Mobile technology has immense potential for addressing pandemics like COVID-19, as it gives us access to big data in terms of volume, velocity, veracity, and variety. These data are particularly relevant to understand and mitigate the spread of pandemics such as COVID-19. In order for such intensive and potentially intrusive data collection measures to succeed, we need a cohesive societal effort with full buy-in from citizens and their representatives. This article outlines an evidence-based global digital citizen science policy that provides the theoretical and methodological foundation for ethically sourcing big data from citizens to tackle pandemics such as COVID-19.

(J Med Internet Res 2020;22(5):e19357) doi:10.2196/19357

KEYWORDS

COVID-19; pandemic; citizen science; smartphones; population health; mHealth; eHealth; big data; virus; infectious diseases; public health; digital epidemiology

A Pandemic in a Digital World

When our society has faced existential crises in the past, we have banded together to overcome the challenge. The coronavirus disease (COVID-19) pandemic is one such threat that requires not only a cohesive effort, but also enormous trust to follow public health guidelines, maintain social distance, and share basic necessities [1,2]. Are democratic societies with civil liberties capable of doing this?

The answer to this question is perhaps right in our pockets. The internet connects us all, and with more than 3 billion devices currently in circulation worldwide [3], if there is one ubiquitous tool that leverages the power of the internet like no other, it is the smartphone. The idea of smartphones being one of the primary solutions to this global problem might seem far-fetched until we unpack its potential. Smartphones provide us with the ability to directly reach and engage with a significant proportion of the world's population in near real time, which has immense potential for addressing COVID-19 outbreaks via rapid detection. Moreover, smartphones have the capacity to provide big data via sensors, such as global positioning systems [4].

Smartphones can provide data about population movement patterns [5], which are extremely pertinent to not only understand and implement social distancing and isolation measures, but also to develop predictive epidemiological modeling of virus spread. In addition, smartphones can facilitate real time interventions to modify behavior and link people with health care services—aspects that can be used to manage both the physical and mental health effects of COVID-19. However, for such intensive and intrusive data collection measures to succeed, we need buy-in from citizens.

The policies and responses of governments around the world have varied in speed and intensity [6], but what unites them is the evidence that COVID-19 is highly contagious. The reality

is that the success of government policies to detect, contain, and minimize the spread of COVID-19 lies beyond the health care systems that are currently barely coping with the ever-increasing growth of COVID-19 positive cases [7]. The success of government policies ultimately depends on the willingness of citizens to follow public health guidelines and abide by laws restricting free movement, which is a challenge in free societies where citizens in one country might abide by regulations better than citizens in another country [8,9]. Thus, whether it is the ability to leverage ubiquitous digital tools such as smartphones, or whether it is the success of government policies to manage and minimize the COVID-19 outbreaks, implementation of these strategies ultimately depends on citizen engagement. The key to building public trust is to link citizen science to citizen engagement.

The Digital Citizen Science Policy Solution

Citizen science is a participatory approach that can range from contributory and collaborative methods (data collection and analysis) to cocreation of knowledge (conceptualization and knowledge translation). It can pave the way for increased citizen engagement and crowdsourcing of big data during this crisis. Citizen science has become increasingly interdisciplinary over the past couple of decades [10], which has implications for population health science—a field of science that plays a key role in addressing broad health inequities [11]. Moreover, with the increasing power of citizens to affect change, citizen science is earning a place in national science policies of countries such as the United States and Australia by complementing the efforts of governments and health professionals [10].

Citizen science offers an opportunity to transform population health science by engaging a larger proportion of the population in data collection to bring citizen perspectives closer to traditional decision-making processes [10]. However, there is currently no coherent citizen science policy to tackle the COVID-19 pandemic. The success of citizen science depends on innovative mobile health (mHealth) apps. The ultimate purpose of mHealth and citizen science, particularly from a disease risk management point of view, is to enable digital epidemiological modeling to prevent, detect, and manage the current wave of the COVID-19 pandemic, as well as to predict and prepare for subsequent waves. One of the biggest challenges in the current COVID-19 pandemic has been the inability to effectively track the virus and population movements that facilitate the spread of the disease.

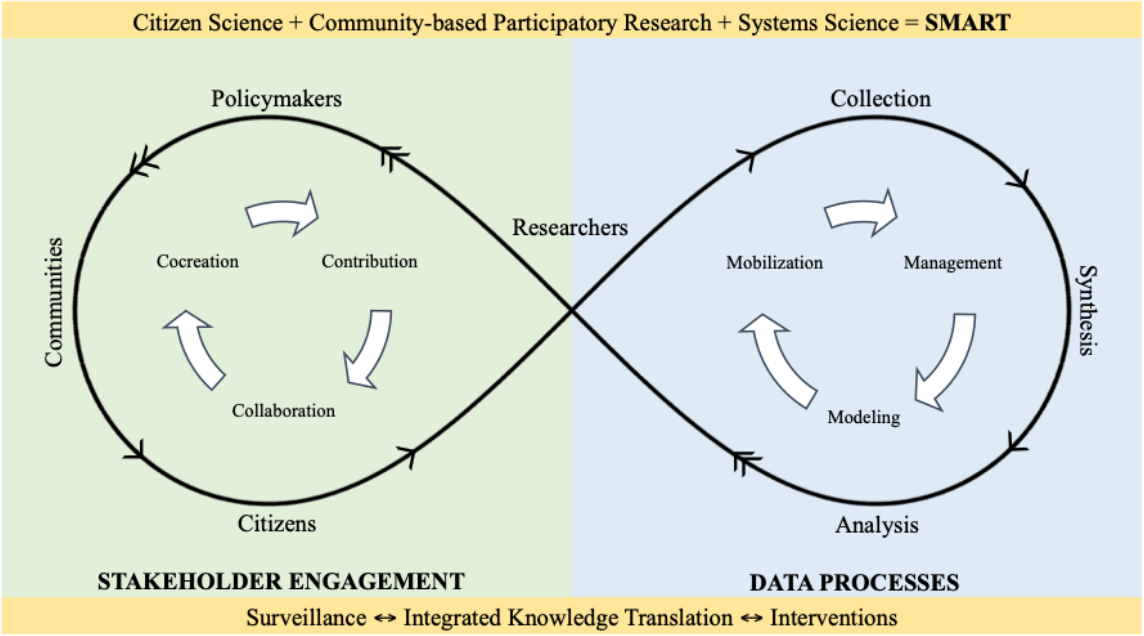
Digital epidemiology is a novel field of science that has been growing rapidly in the past few years. It has been fueled by the increasing availability of data and computing power, as well as by breakthroughs in data analytics [12,13]. The digital health revolution has transformed the collection and analysis of electronic health records, as well as physiological and behavioral measurements at the individual level. However, our health care systems are primed for benefiting downstream service providers for disease management rather than promoting upstream policies to prevent disease development. Another concern is the lack of ethical engagement in digital health, where the power resides predominantly with researchers and providers.

These challenges point toward the need for policy that addresses the intersection of citizen science, innovation, and health to facilitate mHealth and digital citizen science platforms that enable ethical surveillance, integrated knowledge translation, and real time interventions [14,15]. This combination is critical to address population health crises such as the COVID-19 pandemic. The creation of mHealth and digital citizen science platforms can enable:

- Implementation of ethical real time surveillance to assess COVID-19 community risk by subjective and objective (sensor-based) longitudinal data gathering.
- Implementation of near real time integrated knowledge based on the data provided by crowdsourced citizens to increase awareness of health-promoting practices.
- Implementation of evidence-based real time communication that not only link citizens to health care services but also ensures safe social distancing and psychological support.
- Development of decision-making dashboards at the back end, which enhances real time information sharing and data analytics that would inform community decision making to mitigate risk and ensure productivity during health crises.
- Addressing future pandemics by developing models of risk mitigation, which would not only prepare communities, but also enhance health system capacity for future communicable diseases.

The development of such platforms requires robust theoretical underpinnings that provide stakeholders a clear pathway to decision making. One such pathway is the SMART Framework [11]. It integrates citizen science, community-based participatory research, and systems science through ubiquitous tools to conduct population health research in the digital age (Figure 1). A key component of this framework is the necessity to repurpose citizen-owned ubiquitous communication devices (ie, smartphones) that have revolutionized the ability to sense, share, and link big data. This potential for repurposing smartphones is magnified in the current COVID-19 crises because smartphones have the reach to create equity by empowering disenfranchised citizens, and smartphone-based apps have the capacity to source big data to inform policies through the voice of the citizens.

Figure 1. The SMART Framework.



Operationalizing Digital Citizen Science Policy

The flexibility of mHealth and digital citizen science platforms to scale-up opens enormous opportunities to operationalize policy across jurisdictions, whether within a single country or across a cohort of nations. For instance, from a Canadian perspective, health care implementation is the mandate of provinces and territories [16], so natural authorities to introduce digital citizen science policies would be the provinces and territories, where resources are obtained from health care budgets. The federal governments can play a larger role in mandating such policies, which will eventually play a part in enrolling entire populations, a critical factor in the success of mHealth and digital citizen science platforms.

However, digital citizen science policies can exist outside traditional health care systems, as this approach takes a population health lens of health in all policies, which has its own advantages of systems integration [17]. Because citizen science platforms can exist outside health care systems, they do not have to exist within government ecosystems.

In fact, mHealth and digital citizen science platforms have traditionally been used for noncommunicable disease monitoring and interventions, which are implemented outside health care systems by academic research units [14,15]. As the same approach can be used to tackle pandemics like COVID-19, there are opportunities to commercialize mHealth and digital citizen science platforms that will break the silos of jurisdictional authorities. One scenario is that big technology companies, which already have access to a significant proportion of big data collected passively from smartphones across the world, will become major players in disease management [18]. The pros and cons of such a scenario need to be debated comprehensively and are beyond the scope of this policy document. However, the opportunity to commercialize and introduce mHealth and digital citizen science platforms allows groups at risk, such as Indigenous communities, to take control and tackle the risk of pandemics such as COVID-19 [19]. Thus, although the operationalization digital citizen science policies can take many shapes and forms, it is apparent that government agencies will lose their grip on digital health if they do not act now [20].

Benefits of Citizen-Focused Digital Health

Apart from the obvious benefits to the health of populations by reducing the risk of communicable diseases like COVID-19, when digital health takes a citizen science approach, it can enable community empowerment by connecting citizens for a common cause, manage misinformation by directly engaging citizens, and inform evidence-based decision making using big data. There are multiple benefits that would result from linking citizen science and citizen engagement.

The social and societal benefits are:

- Community empowerment: citizen science connects citizens with one another, facilitating engagement, knowledge sharing, and problem solving, especially during existential crises.
- Decision making: citizen science can bring citizens closer to traditional decision-making processes by contributing to the greater dialogue around issues that are of the highest concern to society.
- Misinformation management: citizen science can play an important role in countering digital misinformation.

The population health benefits are:

- Prevention, early detection, and policy interventions: this approach would engage citizens in real time to predict and address global population health risks such as the COVID-19 outbreak, as well as link these risks with upstream policy and downstream health care use data to develop preventive public health policies.
- Evaluating population health interventions: a digital citizen science approach provides the ability to evaluate population health interventions in different jurisdictions, which are especially pertinent when different and incohesive municipal, provincial, and federal policies to counter a pandemic such as COVID-19 are being implemented worldwide.
- Real time interventions: the ability to engage with citizens in real time to deploy behavioral interventions that can help manage population health crises such as COVID-19 are critical to implement measures such as social distancing and mental health support for self-isolating populations across the world.

The benefits to decision makers are:

- Development of decision-making dashboards: real time big data can be used to develop decision-making dashboards that incorporate advanced data analytics to inform global policies during pandemics and manage productivity during social distancing, if necessary.
- Assessing environmental effects: using citizen science, we can understand the effect of physical, social, and cultural environments on health behaviors such as social distancing and abiding by public health guidelines, aspects that are essential to flatten the curve of pandemics.
- Identifying and addressing environmental risks: the trajectory of pandemics such as COVID-19 need to be addressed by taking into consideration varied environmental risks in different geographic regions of the world, and citizen scientists could play an important role in enabling global risk management of pandemics.

The economic benefits are:

- Licensing user access: mHealth and digital citizen science platforms can be licensed to external users on a large-scale to enable commercially viable digital epidemiological studies worldwide.
- Developing mHealth smartphone apps: similarly, mHealth and citizen science platforms can be the source of commercially viable front end mHealth smartphone apps that can be developed rapidly to tackle population health crises such as pandemics.
- Commercialization of dashboards: there is the potential to commercialize decision-making dashboards for different jurisdictions from mHealth and digital citizen science platforms to facilitate real time information sharing worldwide.

Anticipating and Addressing Challenges

Data Privacy and Security

One of the biggest challenges to deploying citizen science at any scale is individuals' right to privacy. As smartphone-based citizen scientist data are granular and detailed owing to the leverage of sensors such as global positioning systems, protecting privacy and anonymity of citizens through strong encryption processes must be the highest priority [11,21,22]. Moreover, before any citizen scientist engagement, obtaining informed consent has to be mandatory. Apart from obtaining informed consent, citizen scientists should be provided an option to dropout and delete their own data (ie, embedding of open source features as much as possible). However, to truly conduct ethical surveillance, data co-ownership is essential, where citizen scientists are able to participate in data visualization, analysis, and knowledge translation [9,12,13]. Ethical surveillance is paramount as COVID-19 tracking escalates across the world because the elimination of ethical data collection now can potentially open doors for invasive surveillance in the future [23]. Apart from equitable engagement with citizens, ethical surveillance also depends on advanced encryption and secure server storage processes, which abide by jurisdictional laws. These capabilities need to be embedded into digital platforms for citizen science to ensure the data that can lead to citizen identification is secured with rigorous ethical and analytical protocols. More importantly, it should be made clear to citizens that anonymity is the guiding principle behind all data privacy and security procedures.

Data Validity and Linkages

Data validity eventually depends on successful and consistent citizen science engagement with digital tools that facilitate secure data collection, synthesis, analyses, and dissemination [24]. A significant factor in citizen scientist engagement would be intuitive front-end human-computer interfaces such as easy to use smartphone apps, and back-end databases that encapsulate data management, modeling, and mobilization. An important feature in ensuring the success of digital citizen science platforms is citizen contribution to the development of both front-end apps and back-end decision-making dashboards. Open-source back-end data management systems [25] can facilitate collaboration across sectors and disciplines via linkages with other open-source databases to maximize the potential of citizen science [26-29]. Nevertheless, the future of citizen science platforms in digital health is dependent on linking upstream administrative, policy, and behavioral citizen data that potentially exist outside the health care system with downstream health care use data. Ultimately, these data linkages require not only legislative action but also digital and logistical infrastructure.

Internet Inequity

The most systemic barrier to digital citizen science is internet inequity. Internet inequity is defined as differential internet access based on wealth, location (urban, rural, or remote), gender, age, or ethnicity [11]. Thus, a risk of the digital citizen science approach is the potential to widen existing health disparities by excluding vulnerable populations who do not have access to the internet [30]. With respect to smartphone dependence, evidence indicates that minority groups and younger, lower income, and less-educated users are more likely to be dependent on smartphones to access the internet [31], an inequity that could be turned into an opportunity via mHealth

and digital citizen science approaches. The digital divide is a complex phenomenon, with varying internet bandwidth across different low-, middle-, and high-income countries. However, there seems to be a connection between bandwidth divide and income divide worldwide [31]. Ultimately, as the United Nations has declared that access to the internet is a human right [32], policy makers have a moral responsibility to address internet inequity, and crises such as the COVID-19 pandemic should provide the impetus for internet equity.

Legitimization and Citizen Scientist Compliance

As the COVID-19 experience has demonstrated, pandemics present public health, economic, and social challenges on a global scale unlike anything else we have witnessed. Thus, tackling pandemics will require transformational change with significant support in terms of personnel training, funding, and time, and increases in transparency to balance the power between decision makers, researchers, and citizens. The COVID-19 challenge provides the prospect of overturning the traditional neglect that citizen science has endured to offer new insights for solving population health crises [33]. However, the ultimate success of citizen science is dependent on effective engagement. The challenges of citizen scientist recruitment, retention, and compliance need to be addressed by a combination of logistical, technological, and methodological solutions before, during, and after data collection.

The integration of citizen science with community-based participatory research can aid citizen engagement and empowerment [12]. Another important factor is developing strategies that are specific to different cohorts, demographic groups, and jurisdictions by taking into consideration historical, cultural, and sociopolitical contexts of populations. Nevertheless, the big question is, regardless of the need for a cohesive societal effort to overcome pandemics, why will citizens ultimately comply? Transformational change requires radical ideas, and if we want citizens to trust institutions, perhaps there is a need to provide citizens with incentives for ethical surveillance. In other words, pay people for sharing data that will enable us to understand the progression of pandemics, and enforce the stringent public health measures necessary to flatten their curve. We cannot secretly monitor citizens; at the same time, we are in uncharted waters when it comes to health and economic consequences of this pandemic. Thus, it is important to seriously consider incentivizing citizens to share data, whether it is through monetization of citizen scientist data, providing tax credits, or other in-kind contributions that would not only ensure equitable participation of citizens but also aid economic equity.

Conclusion

The COVID-19 pandemic is an extremely complex existential threat that requires cohesive societal effort to address health system inefficiencies and to overcome gaps in real time data analytics. When our society has faced existential crises in the past, we have banded together by using the technology at hand to overcome them. Pandemics such as COVID-19, although extremely challenging, offer us an opportunity to innovate and consider options that would normally be ignored due to lack of vision, resources, and coordination. The answer to the current crisis could be right in our pockets. With more than 3 billion smartphones currently in circulation worldwide, they provide us with the ability to directly reach and engage with a significant proportion of the world's population in near real time. However, for such intensive and potentially intrusive data collection measures to succeed, we need a cohesive societal effort with significant buy-in from citizens. A global policy for digital citizen science that facilitates citizen engagement across the world via sophisticated digital epidemiological platforms is a viable solution to not only overcome the COVID-19 outbreak but also better anticipate, prepare, and tackle future pandemics with the rapid response that is necessary in our globalized world.

Acknowledgments

The author acknowledges the Saskatchewan Health Research Foundation and the Canadian Institutes of Health Research for their support to the Digital Epidemiology and Population Health Laboratory (DEPtH Lab) and the SMART Platform. The evidence cited in this article is a result of research conducted using the Saskatchewan Health Research Foundation's Establishment Grant (3779) and the Canadian Institutes of Health Research's Project Grant (#153226). The author also acknowledges the support of Dale Eisler, senior policy fellow, Johnson Shoyama Graduate School of Public Policy.

Conflicts of Interest

None declared.

References

1. Cohut M. Medical News Today. 2020 Mar 24. Why social distancing is key in containing the new coronavirus URL: https://www.medicalnewstoday.com/articles/why-social-distancing-is-key-in-containing-the-new-coronavirus [accessed 2020-04-09]
2. Government of Canada. 2020. Community-based measures to mitigate the spread of coronavirus disease (COVID-19) in Canada URL: https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection/health-professionals/public-health-measures-mitigate-covid-19.html [accessed 2020-04-09]
3. Statista. Smartphone users worldwide 2020 URL: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/ [accessed 2020-01-16]

4. Kang W, Nam S, Han Y, Lee S. Improved heading estimation for smartphone-based indoor positioning systems. 2012 Nov 29 Presented at: IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC); 2012; Sydney, NSW, Australia. [doi: 10.1109/pimrc.2012.6362768]
5. del Rosario M, Redmond S, Lovell N. Tracking the evolution of smartphone sensing for monitoring human movement. Sensors 2015 Jul 31;15(8):18901-18933. [doi: 10.3390/s150818901] [Medline: 26263998]
6. International Monetary Fund. 2020. Policy responses to COVID-19 URL: https://www.imf.org/en/Topics/imf-and-covid19/Policy-Responses-to-COVID-19 [accessed 2020-04-09]
7. Feuer W. CNBC. 2020 Mar 20. WHO officials warn health systems are ‘collapsing’ under coronavirus: ‘this isn’t just a bad flu season’ URL: https://www.cnbc.com/2020/03/20/coronavirus-who-says-health-systems-collapsing-this-isnt-just-a-bad-flu-season.html [accessed 2020-04-09]
8. India TV. 2020 May 04. Social distancing what? Long queues, chaos as liquor shops reopen across country URL: https://www.indiatvnews.com/business/news-liquor-shops-reopen-india-lockdown-3-ground-report-photos-614111 [accessed 2020-05-05]
9. Aleem Z. Vox. 2020 May 03. South Korea has brought new coronavirus cases under control. It’s taking steps to reopen public life URL: https://www.vox.com/2020/5/3/21245594/south-korea-coronavirus-social-distancing-relax [accessed 2020-05-05]
10. Rowbotham S, McKinnon M, Leach J, Lamberts R, Hawe P. Does citizen science have the capacity to transform population health science? Crit Public Health 2017 Nov 09;29(1):118-128. [doi: 10.1080/09581596.2017.1395393]
11. Katapally TR. The SMART Framework: Integration of Citizen Science, Community-Based Participatory Research, and Systems Science for Population Health Science in the Digital Age. JMIR Mhealth Uhealth 2019 Aug 30;7(8):e14056 [FREE Full text] [doi: 10.2196/14056] [Medline: 31471963]
12. Salathé M. Digital epidemiology: what is it, and where is it going? Life Sci Soc Policy 2018 Jan 04;14(1):1 [FREE Full text] [doi: 10.1186/s40504-017-0065-7] [Medline: 29302758]
13. Snyder M, Zhou W. Big data and health. Lancet Digital Health 2019 Oct;1(6):e252-e254. [doi: 10.1016/s2589-7500(19)30109-8]
14. SMART: A Mobile Health and Citizen Science Platform. 2019. URL: https://www.smartstudysask.com/ [accessed 2020-04-06]
15. Katapally TR, Bhawra J, Leatherdale ST, Ferguson L, Longo J, Rainham D, et al. The SMART study, a mobile health and citizen science methodological platform for active living surveillance, integrated knowledge translation, and policy interventions: longitudinal study. JMIR Public Health Surveill 2018 Mar 27;4(1):e31 [FREE Full text] [doi: 10.2196/publichealth.8953] [Medline: 29588267]
16. Vayda E, Deber RB. The Canadian health care system: an overview. Soc Sci Med 1984 Jan;18(3):191-197. [doi: 10.1016/0277-9536(84)90079-0]
17. World Health Organization. 2010. Implementing health in all policies URL: https://www.who.int/sdhconference/resources/implementinghiapadel-sahealth-100622.pdf [accessed 2020-04-09]
18. Kelion L. BBC. 2020 Apr 10. Coronavirus: Apple and Google team up to contact trace Covid-19 URL: https://www.bbc.com/news/technology-52246319
19. Fuller J, Hermeston W, Passey M, Fallon T, Muyambi K. Acceptability of participatory social network analysis for problem-solving in Australian Aboriginal health service partnerships. BMC Health Serv Res 2012 Jun 10;12:152. [doi: 10.1186/1472-6963-12-152] [Medline: 22682504]
20. Baharudin H. The Straits Time. 2020 Mar 23. Coronavirus: S'pore Government to make its contact-tracing app freely available to developers worldwide URL: https://www.straitstimes.com/singapore/coronavirus-spore-government-to-make-its-contact-tracing-app-freely-available-to [accessed 2020-04-09]
21. Denecke K. An ethical assessment model for digital disease detection technologies. Life Sci Soc Policy 2017 Sep 20;13(1):16 [FREE Full text] [doi: 10.1186/s40504-017-0062-x] [Medline: 28929347]
22. Kostkova P. Disease surveillance data sharing for public health: the next ethical frontiers. Life Sci Soc Policy 2018 Jul 04;14(1):16 [FREE Full text] [doi: 10.1186/s40504-018-0078-x] [Medline: 29971516]
23. Singer N, Sang-Hun C. As coronavirus surveillance escalates, personal privacy plummets. The New York Times 2020.
24. Crain R, Cooper C, Dickinson JL. Citizen science: a tool for integrating studies of human and natural systems. Annu Rev Environ Resour 2014 Oct 17;39(1):641-665. [doi: 10.1146/annurev-environ-030713-154609]
25. Alabri A, Hunter J. Enhancing the quality and trust of citizen science data. In: Proceedings of the Sixth International Conference on e-Science. 2011 Jan 20 Presented at: eScience'11; December 7-10, 2010; Brisbane, QLD, Australia. [doi: 10.1109/escience.2010.33]
26. Fink D, Damoulas T, Dave J. Adaptive spatio-temporal exploratory models: hemisphere-wide species distributions from massively crowdsourced eBird data. In: Proceedings of the Twenty-seventh AAAI Conference on Artificial Intelligence. 2013 Presented at: AAAI'13; July 14-18, 2013; Bellevue, Washington, USA p. 1284-1290 URL: https://www.semanticscholar.org/paper/Adaptive-Spatio-Temporal-Exploratory-Models%3A-from-Fink-Damoulas/f8856ff340a3e029041377bfe5db0db73659d8d6

27. Tulloch AI, Possingham HP, Joseph LN, Szabo J, Martin TG. Realising the full potential of citizen science monitoring programs. Biol Conserv 2013 Sep;165:128-138. [doi: 10.1016/j.biocon.2013.05.025]
28. Hampton SE, Strasser CA, Tewksbury JJ, Gram WK, Budden AE, Batcheller AL, et al. Big data and the future of ecology. Front Ecol Environ 2013 Apr;11(3):156-162. [doi: 10.1890/120103]
29. Worthington J, Silvertown J, Cook L, Cameron R, Dodd M, Greenwood R. Evolution MegaLab: a case study in citizen science methods. Methods Ecol Evol 2012;3(2):a. [doi: 10.1111/j.2041-210x.2011.00164.x]
30. Zhang X, Pérez-Stable EJ, Bourne PE, Peprah E, Duru OK, Breen N, et al. Big data science: opportunities and challenges to address minority health and health disparities in the 21st century. Ethn Dis 2017;27(2):95-106 [FREE Full text] [doi: 10.18865/ed.27.2.95] [Medline: 28439179]
31. Hilbert M. The bad news is that the digital access divide is here to stay: domestically installed bandwidths among 172 countries for 1986–2014. Telecommunications Policy 2016 Jun;40(6):567-581. [doi: 10.1016/j.telpol.2016.01.006]
32. Office of the United Nations High Commissioner for Human Rights. 2011. Report of the special rapporteur on the promotion and protection of the right to freedom of opinion and expression, Frank La Rue URL: https://www2.ohchr.org/english/bodies/hrcouncil/docs/17session/A.HRC.17.27_en.pdf [accessed 2019-07-12]
33. Thompson B, Molina Y, Viswanath K, Warnecke R, Prelip ML. Strategies to empower communities to reduce health disparities. Health Aff 2016 Aug 01;35(8):1424-1428 [FREE Full text] [doi: 10.1377/hlthaff.2015.1364] [Medline: 27503967]

Abbreviations

COVID-19: coronavirus disease
mHealth: mobile health

Edited by G Eysenbach; submitted 14.04.20; peer-reviewed by C Fincham, D Gunasekeran, C Cleland, S Pesälä; comments to author 04.05.20; revised version received 05.05.20; accepted 14.05.20; published 26.05.20.

Please cite as:
Katapally TR
A Global Digital Citizen Science Policy to Tackle Pandemics Like COVID-19
J Med Internet Res 2020;22(5):e19357
URL: http://www.jmir.org/2020/5/e19357/
doi:10.2196/19357
PMID:32408267

Original Paper

Public Engagement and Government Responsiveness in the Communications About COVID-19 During the Early Epidemic Stage in China: Infodemiology Study on Social Media Data

Qiuyan Liao[1*], PhD; Jiehu Yuan[1*], MA; Meihong Dong[1*], MPH; Lin Yang[2], PhD; Richard Fielding[1], PhD; Wendy Wing Tak Lam[1,3], PhD

[1]School of Public Health, The University of Hong Kong, Hong Kong

[2]School of Nursing, The Hong Kong Polytechnic University, Hong Kong

[3]School of Nursing, The University of Hong Kong, Hong Kong

[*]these authors contributed equally

Corresponding Author:
Qiuyan Liao, PhD
School of Public Health
The University of Hong Kong
2/F, Patrick Manson Building (North Wing)
7 Sassoon Road, Pokfulam
Hong Kong
Phone: 852 39179289
Email: qyliao11@hku.hk

Abstract

Background: Effective risk communication about the outbreak of a newly emerging infectious disease in the early stage is critical for managing public anxiety and promoting behavioral compliance. China has experienced the unprecedented epidemic of the coronavirus disease (COVID-19) in an era when social media has fundamentally transformed information production and consumption patterns.

Objective: This study examined public engagement and government responsiveness in the communications about COVID-19 during the early epidemic stage based on an analysis of data from Sina Weibo, a major social media platform in China.

Methods: Weibo data relevant to COVID-19 from December 1, 2019, to January 31, 2020, were retrieved. Engagement data (likes, comments, shares, and followers) of posts from government agency accounts were extracted to evaluate public engagement with government posts online. Content analyses were conducted for a random subset of 644 posts from personal accounts of individuals, and 273 posts from 10 relatively more active government agency accounts and the National Health Commission of China to identify major thematic contents in online discussions. Latent class analysis further explored main content patterns, and chi-square for trend examined how proportions of main content patterns changed by time within the study time frame.

Results: The public response to COVID-19 seemed to follow the spread of the disease and government actions but was earlier for Weibo than the government. Online users generally had low engagement with posts relevant to COVID-19 from government agency accounts. The common content patterns identified in personal and government posts included sharing epidemic situations; general knowledge of the new disease; and policies, guidelines, and official actions. However, personal posts were more likely to show empathy to affected people (χ^2_1=13.3, P<.001), attribute blame to other individuals or government (χ^2_1=28.9, P<.001), and express worry about the epidemic (χ^2_1=32.1, P<.001), while government posts were more likely to share instrumental support (χ^2_1=32.5, P<.001) and praise people or organizations (χ^2_1=8.7, P=.003). As the epidemic evolved, sharing situation updates (for trend, χ^2_1=19.7, P<.001) and policies, guidelines, and official actions (for trend, χ^2_1=15.3, P<.001) became less frequent in personal posts but remained stable or increased significantly in government posts. Moreover, as the epidemic evolved, showing empathy and attributing blame (for trend, χ^2_1=25.3, P<.001) became more frequent in personal posts, corresponding to a slight increase in sharing instrumental support, praising, and empathizing in government posts (for trend, χ^2_1=9.0, P=.003).

Conclusions: The government should closely monitor social media data to improve the timing of communications about an epidemic. As the epidemic evolves, merely sharing situation updates and policies may be insufficient to capture public interest

in the messages. The government may adopt a more empathic communication style as more people are affected by the disease to address public concerns.

(J Med Internet Res 2020;22(5):e18796) doi:10.2196/18796

KEYWORDS

risk communication; social media; epidemic; COVID-19; pandemic; outbreak; infectious disease; content analysis

Introduction

Background

On December 31, 2019, a cluster of pneumonia cases of unknown etiology were first reported in Wuhan, the capital city of Hubei Province, China [1]. The causative pathogen was soon identified as a novel coronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [2] and the disease caused by SARS-CoV-2 was termed the coronavirus disease (COVID-19) [3]. Epidemiological investigation of the first 41 laboratory-confirmed human cases revealed that most had a history of visiting a seafood wholesale market (Huanan Market) in Wuhan where live wild animals were also sold for human consumption [4], but later, human-to-human transmission was confirmed, as cases without a history of visiting the market increased dramatically [5-7]. The outbreak of COVID-19 in Wuhan rapidly evolved into a severe pneumonia epidemic nationally and later worldwide. As of May 13, 2020, a total of 84,458 confirmed human cases of COVID-19 including 4644 deaths in China were reported to the World Health Organization [8]. Worldwide, COVID-19 has been a pandemic affecting more than 200 countries or territories with a total of 4,170,424 human cases of COVID-19 including 287,399 deaths as of May 13, 2020.

Communication About Outbreaks

The outbreak of a newly emerging respiratory infectious disease usually puts individuals at a high risk of infection and constitutes a highly uncertain situation that changes rapidly, threatening serious potential loss and prompting considerable psychological distress [9,10]. Feelings of uncertainty provoke great public anxiety, which if not properly addressed can develop into public panic and herd behaviors that may harm social order and population health [11,12]. Effective outbreak communication, particularly at the early stage, becomes critically important for dealing with excessive public fear, promoting risk awareness, empowering the public in taking protective actions, and gaining public confidence and trust [11,13,14]. Conceptual models for guiding outbreak communications have been developed since the 2003 outbreak of severe acute respiratory syndrome (SARS) [14-17] and used in empirical research for examining communication practices during the 2009 influenza A/H1N1 pandemic [14], the 2013-2016 Ebola outbreak in West Africa [18], and most recently the outbreak of COVID-19 [19]. These communication models and empirical research indicate that effective outbreak communication should be prompt and transparent, dynamic as the situation evolves to meet changes in public needs, relevant and able to engage the community, and empathic and caring to address public emotional distress [14-19].

Social Media as a Platform for Outbreak Communication

The high penetration of internet use and rapid development of information and communication technologies have made the internet an increasingly important health information source worldwide [20-23]. Reading, commenting, sharing, and seeking health information from social media, particularly through a mobile device, has become an increasingly important pattern of health information consumption in China [20,24,25]. During an epidemic, social media can facilitate the spread of epidemic awareness, attitudes toward control and preventive measures, emotional responses and behaviors, as well as misinformation and rumors in the public through online interactivity [26-30]. As the epidemic evolves, this may facilitate homogeneous mental representations of the epidemic, leading to collective behavioral responses [31]. In China, Sina Weibo (the Chinese version of Twitter) is one of the most popular platforms that attracted 486 million active monthly users in 2019 [32], most of whom accessed their user accounts through a mobile device [24]. Its microblogging function allows users to create and share short messages in a multimedia format, and other users can "share," "like," "comment," and "follow" the initial posts. Numerous government agencies in China also make use of Weibo to communicate with the public. As of June 2019, there were a total of 139,270 verified government microblogs in Weibo [24].

With the proliferation of internet and social media used as health information sources, infodemiology (or infoveillance), the study (or surveillance) of "distribution and determinants of information," in the internet or a population for the aim of guiding policy making and public health interventions [33] has been commonly used in the case of infectious disease outbreaks. Among various applications of infodemiology or infoveillance methods for social media data about infectious disease outbreaks, tracking information (concept) prevalence data [27,34-38] and qualitatively analyzing and categorizing contents of the social media data [19,27,36-41] are believed to be able to provide important insights into outbreak communications. However, existing studies mainly focused on tracking specific concepts or information such as blame [27], misinformation [34], stigma [30], specific keywords and sentiment [35], and organization trust and managing uncertainty [39], possibly due to specific research interests or using machine-aided analysis which does not allow flexible content analysis [42]. Moreover, existing studies mainly focused on one side of the outbreak communication, either the public response or response of health authorities [19,27,34-41]. This only provides partial understanding about the interactivity between the public and health authorities, whereas two-way communication is believed to be crucial for effective outbreak communication [40].

Although a recent infoveillance study on Weibo data about COVID-19 provided some qualitative descriptions about the potential interactions between the public and government online about COVID-19 by time, the study did not quantify public engagement and government responsiveness regarding COVID-19 and how these changed by time as the pandemic unfolded [38]. In addition, this study's qualitative results seemed to lack a clear structure for understanding public perceptions and emotions [38]. The study conducted by Chew and Eysenbach [37] provides more comprehensive methods for guiding coding of social media data related to discussion topics, emotions, and online behaviors as well as quantifying and tracking the changes of these contents as the epidemic unfolds.

Study Objectives

The outbreak of COVID-19 was an unprecedented epidemic that China experienced for the first time in this digital era. Based on this study's literature review on important principles of effective outbreak communication and the knowledge gaps of current infodemiology or infoveillance studies examining outbreak communication using social media data, this study aimed to make use of the Weibo data about COVID-19 from December 2019 to January 2020 in China to answer the following research questions:

1. How did the public respond to the outbreak of COVID-19 as cases of COVID-19 increased and increasingly stringent containment measures were implemented, and how quickly could the government respond to public discussions about COVID-19?
2. To what extent could government messages about COVID-19 engage the general online users?
3. What contents did the public discuss online, and how did these contents change as the epidemic evolved? To what extent could the government respond to the temporal change of public discussions online?

Methods

Data Extraction

Four keywords in Chinese characters were used to capture data relevant to COVID-19 from Weibo from December 1, 2019, to January 31, 2020: "Wuhan pneumonia," "novel coronavirus," "novel coronavirus pneumonia," and "novel pneumonia." Data were retrieved using the built-in Weibo searching function and were subsequently screened by the Python Web Crawler, a tool that has been demonstrated to be efficient in identifying the most relevant posts that contain the set keywords [43]. A total of 1,028,204 relevant posts were initially retrieved. We also tried the keywords "unknown pneumonia" and "SARS" in Chinese characters to capture Weibo data from December 1, 2019, to January 9, 2020, when the etiology of COVID-19 was not yet confirmed [2]. Since "Wuhan pneumonia" is a less specific term for COVID-19 before official announcement of unknown pneumonia in Wuhan on January 9, 2020, we manually checked the relevant posts by January 9, 2020, in the database. This excluded 466 from 469 posts on December 30, 2019, but this term successfully detected over 99% of the posts relevant to COVID-19 in the subsequent days. The final database included a total of 1,027,738 posts comprising of 914,247 (89.0%) posts from personal accounts of the public, 45,398 (4.4%) posts from accounts of government agencies, and 67,746 (6.6%) posts from accounts of media and commercial agencies. Accounts of government, media, and commercial agencies were verified by Weibo to be "official" at registration by submitting relevant documents of their organizations for verification. Each post record comprises account name, contents, post time, engagement data of each post including numbers of likes (ie, showing confirmation or agreement with the post contents), comments, shares, and followers.

Engagement Analysis

This study evaluated how much government posts can engage online users in the communications about COVID-19 by calculating the engagement index using engagement data of the posts delivered by government agency accounts [44,45]. The engagement data comprising likes, shares, and comments on the posts from each government account and the number of followers of these accounts were first extracted. These engagement data were then used to calculate the three metrics of social media engagement: popularity (likes per post and per 1000 followers), commitment (comments per post and per 1000 followers), and virality (shares per post and per 1000 followers); all three metrics were subsequently aggregated to generate the overall engagement index. Based on the engagement index, we identified the top five most active government agencies in both the health and nonhealth sectors. When ranking engagement, we specifically excluded any government account that delivered fewer than the average number of posts generated by all government accounts combined during the study period because these government accounts could rank high based on engagement index but were considered inactive in interacting with online audiences. We used mean but not the median number of posts as the cut-off because over 50% of government agencies only had 1 post during the study period. In addition, since the National Health Commission (NHC) of China is regarded as the lead agency in coordinating the national effort to combat the COVID-19 outbreak in China, its engagement data were included in comparison with other government agencies despite that its engagement index was not ranked in the top five.

Content Analysis

There is currently no consensus about how to best sample social media data for content analysis [42], but random sampling has been commonly used and seems to be suitable for social media data [46]. However, for random sampling, sample sizes differ a lot across studies due to different study purposes, duration of study period, resources, and whether data were coded automatically or manually [42]. Manual coding can generate richer information by accommodating new codes emerging during data analysis, but the sample size must be kept at a manageable level to avoid fatigue and improve accuracy in coding. Since we were interested in temporal changes in the discussion contents about COVID-19 among general online users, we focused on the personal account posts for content analysis. We first excluded personal account posts that lacked all engagement (no likes, comments, and shares) because these posts may have captured little attention and interest of other online users. Thereafter, 20 posts per calendar day between

December 31, 2019, and January 31, 2020, were randomly selected for content analysis. In addition, 4 posts delivered by personal accounts (1 on December 29, 2019, and 3 on December 30, 2019) before the first official announcement of the unknown pneumonia cases in Wuhan on December 31, 2019, were included for content analysis. Therefore, a total of 644 personal account posts were finally included for content analysis.

As a comparison with personal account posts and supplement for understanding government responsiveness in the communications about COVID-19, we also analyzed the contents of all posts from the first 5 government accounts of both the health and nonhealth sectors based on the rank of the engagement index and posts from the NHC in the same period. The first government post relevant to COVID-19 was posted on December 31, 2019. A total of 273 government posts were included for content analysis.

A tentative coding scheme was first developed based on preliminary analysis of a random training subset of 100 posts using open coding by one author (QL) and iteratively refined through independent analysis of another 100 training posts by authors QL, JY, and MD, and discussion from the team. The coding scheme was then used by two authors (JY and MD) to analyze the 274 selected government posts and 644 selected individual posts for final content analysis, each analyzing half of these posts. Although the coding scheme was used, the coders were advised to be open to new codes during the analysis. After both coders finished their part, they mutually checked 10% of the posts from each other's subset to ensure consistency in coding. Finally, one author (QL) double-checked a random subset of 10% of all posts for content analysis to ensure the accuracy and reliability of the coding. Any inconsistencies were solved by going back to relevant data and joint discussions among authors QL, JY, and MD to reach an agreement. Interrater reliability was assessed by calculating Cohen kappa, with a value of 0.6 or above indicating adequate reliability. The set of codes finally generated were then constant compared to develop thematic categories.

Statistical Analysis

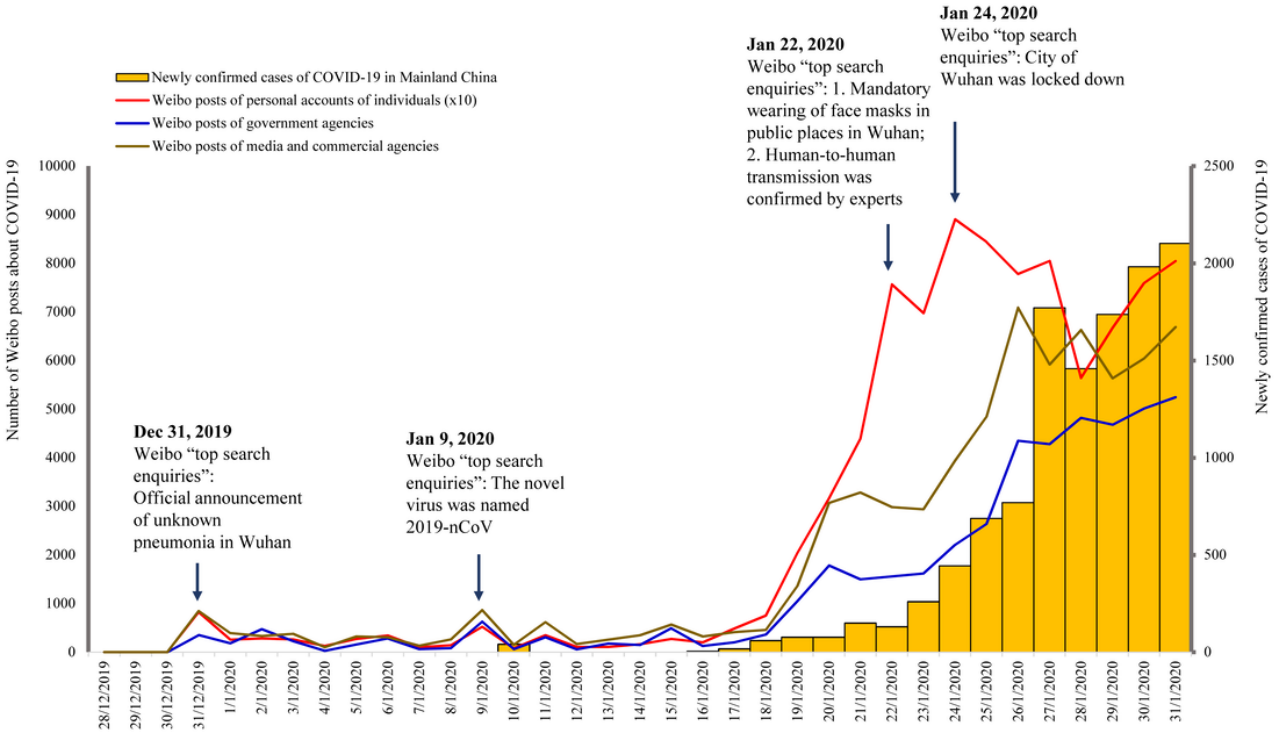
A Pearson chi-square test was used to compare the overall percentages of each thematic category between personal account posts and government agency posts. Latent class analysis (LCA) was used to explore main patterns of post contents by type of account. To conduct the LCA, we first generated a binary-coded variable (1=the specific thematic category is present and 0=the specific thematic category is absent) for each thematic category coded for each post. These variables were then entered into Mplus 7.3 (Muthén and Muthén) for LCA. Major fit indices provided by Mplus including Akaike information criterion (AIC), Bayesian information criterion (BIC), sample size adjusted BIC (aBIC), and the entropy value were used to determine the optimal number of class for the post contents of both personal accounts and government agency accounts. A model with smaller values of AIC, BIC, and aBIC but with a higher entropy value is preferred, but we also consider the interpretability of the model and model parsimony [47]. The LCA would finally determine the content pattern (ie, latent class) of each post. A chi-square for trend was then used to examine the temporal change of the proportion of each content pattern by week by type of account after checking the linearity of the distribution of proportion of each content pattern by week. The temporal trend analysis excluded the 4 personal posts delivered before December 31, 2019. The LCA was conducted using Mplus 7.3, and other statistical analyses were conducted using STATA 15.1 for Windows (StataCorp LLC).

Results

Weibo Activity

We detected information prevalence in relation to the outbreak by plotting daily numbers of Weibo posts by type of accounts with daily numbers of newly confirmed cases of COVID-19 in Figure 1. The Weibo "top search enquires" automatically identified by the built-in function of Weibo based on the number of relevant posts on that day were also marked on observable peak days of Weibo activity (Figure 1). There was a peak on December 31, 2019, when a cluster of unknown pneumonia cases in Wuhan were officially announced for the first time. Another small peak was detected on January 9, 2020, when there was a hot discussion about naming the etiology of unknown pneumonia in Wuhan as the novel coronavirus. Weibo activity increased dramatically starting from January 18, 2020, when daily new confirmed cases of COVID-19 substantially increased. For personal account posts, a third peak was found on January 22, 2020, when the Wuhan government announced a policy of mandatory wearing of face masks in public places and human-to-human transmission of COVID-19 was announced by an expert who had just visited Wuhan for investigations, and the last peak within our study timeframe was detected on January 24, 2020, within 24 hours after Wuhan city was locked down. For posts from government agencies and media and commercial agencies, the third peak was not observed, and the last peak within the study time frame occurred 2 days later. Overall, it appeared that public online reactions followed the spread of the disease and government actions, and responded earlier than the government.

Figure 1. Daily numbers of newly confirmed cases of COVID-19 in Mainland China, daily numbers of Sina Weibo posts relevant to COVID-19 by account, and Weibo "top search enquiries" on peak days from December 2019 to January 2020. COVID-19: coronavirus disease; 2019-nCoV: novel coronavirus.



Public Engagement With Government Messages

Table 1 presents the engagement metrics of the five most active Weibo accounts of government agencies from both the health and nonhealth sectors in the communications about COVID-19 on Weibo. In the health sector, the most active government agencies are the Municipal Health Commission (MHC) in several cities of China, including Wuhan, Zhuhai, Shanghai, and Beijing, and one city-level hospital in Sichuan Province, China. In China, the MHC in different cities are required to report duties to the provincial and national health commissions. In the nonhealth sector, the most active government agency was the Hubei Branch of the Red Cross Society of China, an organization supervised by the Chinese government that is mainly responsible for encouraging donation to support affected people during the epidemic. The remainders are from the system of public security bureaus that are responsible for tracking close contacts of patients with COVID-19 and implementing local policies of traffic restrictions during the epidemic. The engagement analysis shows that, despite being relatively more active compared with other government agency accounts based on the rank of the engagement index, other than Wuhan MHC, all government agencies had low popularity (likes per post per 1000 followers), commitment (comments per post per 1000 followers), and virality (shares per post per 1000 followers). The NHC, despite having attracted a large number of followers, had extremely low popularity, commitment, and virality.

Table 1. Government agencies with the greatest engagement and associated engagement metrices for coronavirus disease communications from December 2019 to January 2020.

Government agencies	Posts, n	Followers, n	Popularity[a], n	Commitment[b], n	Virality[c], n	Engagement index[d]
Top 5 from health sector						
Wuhan MHC[e]	40	58,144	242.73	9.72	2.62	255.07
Zigong No 4 People's Hospital	21	298	42.03	5.27	0.96	48.26
Zhuhai MHC	18	11,844	27.82	3.53	1.18	32.52
Shanghai MHC	11	403,603	25.35	1.84	1.14	28.34
Daxing (in Beijing) MHC	12	65,707	14.41	1.61	1.38	17.41
Top 5 from nonhealth sector						
Hubei Branch of the Red Cross Society of China	21	97,523	78.20	8.70	1.10	88.01
Gaolan People's Procuratorate (in Gansu Province)	22	550	13.06	4.71	1.82	19.59
Suixian People's Procuratorate (in Hubei Province)	59	435	6.97	4.71	2.42	14.10
Longchang Public Security Bureau (in Sichuan Province)	26	8803	12.51	0.73	0.20	13.44
Datong Fire Services Department (in Shanxi Province)	13	2155	8.07	2.75	0.82	11.64
National Health Commission of China	30	5,371,595	7.51	0.33	0.22	8.04

[a]Likes per post per 1000 followers.

[b]Comments per post per 1000 followers.

[c]Shares per post per 1000 followers.

[d]Engagement index = popularity + commitment + virality.

[e]MHC: Municipal Health Commission.

Individual and Government Post Contents

Frequencies and proportions of thematic categories of post contents for personal and government accounts are shown in Table 2, and detailed descriptions of the thematic categories identified in our study can be found in Multimedia Appendix 1.

We noted 3 cyber-support behaviors from personal account posts: *sharing knowledge or information*, *emotional exchange*, and *seeking information* of which only the first 2 were identified in government posts. As is shown in Table 2, for *sharing knowledge and information* in both groups, the most common thematic categories were situation updates of COVID-19 followed by general knowledge about coronavirus pneumonia and advice on preventive measures. Government agency posts were more likely to share information about policies, guidelines, and official actions (χ^2_1=14.5, *P*<.001), and instrumental support (χ^2_1=32.5, *P*<.001), while personal account posts were more likely to share information on public responses to the epidemic (χ^2_1=19.1, *P*<.001). Personal account posts were more likely to be classified as *emotional exchange* (χ^2_1=30.5, *P*<.001) including showing empathy to affected people (χ^2_1=13.3, *P*<.001), attributing blame to people or organizations for malpractice during the epidemic (χ^2_1=28.9, *P*<.001), and expressing worry about the epidemic (χ^2_1=32.1, *P*<.001). The government posts more likely praised people or organizations (χ^2_1=8.7, *P*=.003). The main groups of people praised by both groups were health care workers, while the main people or organizations being blamed in personal account posts included other individuals (eg, individuals who consumed wild animals, breached the infection containment measures, and committed medical violence) and the government (individual government officers or government in general). Regarding *seeking information*, the main information sought was updates about the epidemic situation.

Table 2. Frequency of thematic categories from posts delivered by individual and government accounts.

Thematic categories	Individuals (n=644), n (%)	Government (n=273), n (%)	*P* value[a]
Sharing knowledge/information	567 (88.0)	258 (94.5)	.003
Situation update about COVID-19[b]	287 (44.6)	108 (39.6)	.16
General knowledge about coronavirus pneumonia	206 (32.0)	82 (30.0)	.39
Advice on preventive measures	114 (17.7)	56 (20.5)	.32
Policies, guidelines, and official actions	95 (14.8)	69 (25.3)	<.001
Human-to-human transmission	79 (12.3)	27 (9.9)	.30
Fight against rumors	46 (7.1)	23 (8.4)	.50
Cause of viral emergence	44 (6.8)	14 (5.1)	.33
Public response during the epidemic	43 (6.7)	0 (0.0)	<.001
Instrumental support	13 (2.0)	29 (10.6)	<.001
Infection and illness experience	10 (1.6)	2 (0.7)	—[c]
Seeking social support	10 (1.6)	0 (0.0)	—[c]
Request for information transparency	8 (1.2)	0 (0.0)	—[c]
Reports of scientific research	3 (0.5)	0 (0.0)	—[c]
Seeking close contact	0 (0.0)	2 (0.7)	—[c]
Emotional exchange	321 (49.8)	82 (30.0)	<.001
Showing empathy to or blessing affected people	86 (13.4)	14 (5.1)	<.001
Blaming people or organizations	78 (12.1)	3 (1.1)	<.001
Providing reassurance about risk	73 (11.3)	30 (11.0)	.88
Expressing worry or fear about the risk	70 (10.9)	0 (0.0)	<.001
Praising people or organizations	53 (8.2)	40 (14.7)	.003
Warning about the epidemic	48 (7.5)	11 (4.0)	.05
Seeking information	36 (5.6)	0 (0.0)	<.001

[a]*P* values were calculated using a Pearson chi-square test.

[b]COVID-19: coronavirus disease.

[c]Cell with expected frequency less than 5 and thereby *P* values from the chi-square test were not available.

Compared with post contents of government agencies from the health sector, we found that the four government agencies from the system of public security bureaus were more likely to share information about situation updates (χ^2_2=15.9, *P*<.001), while the Hubei Red Cross Society of China were more likely to post contents about instrumental support (eg, donation of materials or money; χ^2_2=25.3, *P*<.001) and showing empathy to affected people (χ^2_1=25.7, *P*<.001).

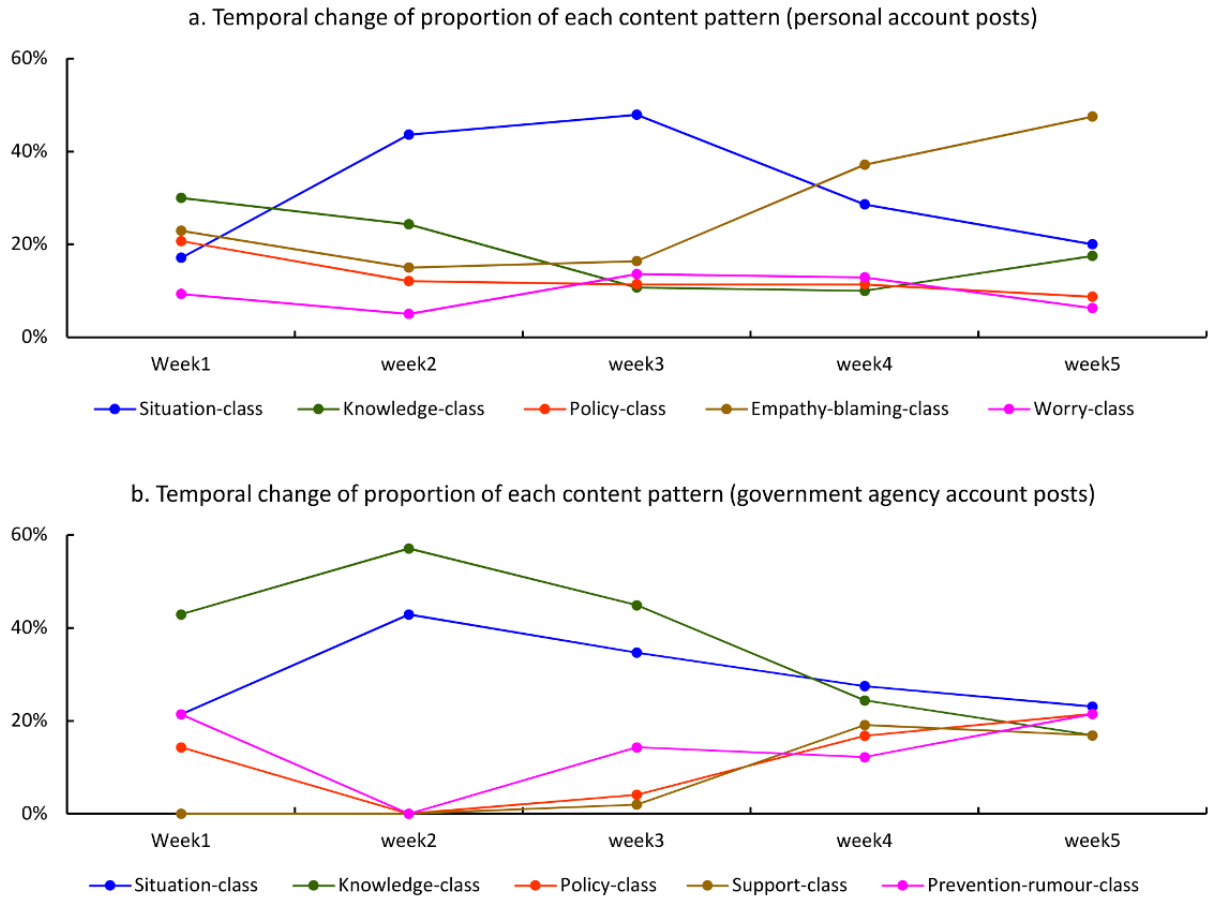
Content Patterns and Temporal Changes

The LCA revealed five main content patterns within both personal posts and government posts (Multimedia Appendix 2). For personal posts excluding the 4 posts before December 31, 2019, the most prevalent content pattern comprised of posts that were mainly sharing situation updates (situation class; n=208/640, 32.5%), followed by those with greater probabilities of showing empathy and blaming (empathy-blaming class; n=166/640, 25.9%); those mainly sharing general knowledge (knowledge class; n=119/640, 18.6%); those mainly sharing policy, guidelines, and official actions (policy class; n=85/640, 13.3%); and those mainly sharing worry about the epidemic (worry class; n=62/640, 9.7%). For the government posts, 3 similar classes as those in the personal posts were found, including the situation class (n=77/273, 28.2%), knowledge class (n=79/273, 28.9%), and policy class (n=40/273, 14.7%; Multimedia Appendix 3). Another 3 classes were different from those of the personal posts, including posts that mainly shared prevention tips and fought against rumors (prevention-rumors class; n=40/273, 14.7%) and those mainly providing instrumental support, praising people, and that had a slightly higher probability of showing empathy (support class; n=37/273, 13.6%; Multimedia Appendix 3).

The temporal change of each content pattern by type of account are shown in Figure 2. The chi-square for trend analyses found that for personal posts, proportions of posts sharing situation updates increased from week 1 (December 31, 2019-January 6, 2020) to week 3 (January 14-20, 2020; for trend, χ^2_1=28.6,

P<.001) but declined thereafter (for trend, χ^2_1=19.7, *P*<.001). Sharing general knowledge (for trend, χ^2_1=15.3, *P*<.001) and policies, guidelines, and official actions (for trend, χ^2_1=15.3, *P*<.001) declined each week, while showing empathy and blaming increased significantly in later weeks (for trend, χ^2_1=25.3, *P*<.001), and worry about the epidemic remained stable and low. For government agency account posts, sharing situation updates and prevention tips, and fighting against rumors remained stable; sharing general knowledge declined each week (for trend, χ^2_1=14.7, *P*<.001), while sharing policies, guidelines, and official actions (for trend, χ^2_1=8.9, *P*=.003), and providing instrumental and emotional empathy and praising people (for trend, χ^2_1=9.0, *P*=.003) increased significantly on the last 2 weeks of the study time frame.

Figure 2. Temporal changes in the proportion of main content patterns by type of account.



We also specifically explored how providing reassurance, a main emotional content in government posts, changed as the epidemic evolved. We found that providing reassurance in government posts was more frequent in the first 3 weeks but declined in the last 2 weeks of the study time frame (for trend, χ^2_1=4.2, *P*=.04)

Discussion

Principal Findings

The Weibo activity in the early stage of the COVID-19 epidemic showed that public reactions seemed to follow the spread of the disease and government actions in China. This again evidences the value of infodemiology or infoveillance studies to understand public response to the disease and government actions during the early epidemic stage despite a concern over the censorship of online information for propaganda purposes in China. There was 1 post blaming people who consumed wild animals for causing “Wuhan pneumonia” that was captured on December 29, 2019. On December 30, 2019, 3 more posts were identified: 2 seeking confirmation about unknown pneumonia cases detected in Hunan Seafood Market in Wuhan and 1 sharing information about unknown pneumonia cases found in the market. All 4 posts were from accounts of individuals who lived in Wuhan. This indicates that before the official announcement of the first cluster of unknown pneumonia in Wuhan on December 31, 2019 [1], relevant information had been spread in the public through interpersonal communication, social media, or other channels. The first government post was delivered on December 31, 2019, 2 days after the first individual post, indicating that, although the Chinese government’s outbreak communication has been more timely and transparent compared with their response to the 2003 SARS outbreak [48], more efforts are required to improve early outbreak communication when uncertainty usually challenges communication. Early posts should be treated as alarms and responded to in a timely

manner rather than being silenced, which could receive harsh criticism and worsen the epidemic control [49].

Our study evidences the use of social media among the Chinese government agencies in the communications about COVID-19 at the early epidemic stage, but the communication remained limited. The engagement analysis indicates that the public generally had low engagement with government agency posts, even those from the most active government agencies. In China, the NHC is expected to play a leading role in risk communication during an epidemic, while provincial and municipal health commissions are expected to report duties and provide epidemic information of their own provinces and cities, respectively, to the NHC. However, although the NHC and provincial health commissions have attracted a large number of followers on social media during the epidemic, compared with municipal health commissions, messages from these high-level health authorities were more disengaged by the general public. The generally low values of the engagement metrics popularity (likes per post per 1000 followers), commitments (comments per post per 1000 followers), and virality (shares per post per 1000 followers) may indicate low levels of interest, use, emotional arousal, or even credibility of the government information among the online audiences [50,51]. This may partly reflect a population inertia effect, where it takes a certain amount of time, or threat, before there is a noticeable change in the bulk practices of a population. The duration of such a period of inertia would be valuable to know.

The content analysis suggests that the Chinese government agencies mainly used social media to "inform" the public about updates of the epidemic situation; knowledge of the coronavirus pneumonia; policies, guidelines, and government actions; and prevention tips, all being the key risk messages included in the official websites of health authorities for communicating about an epidemic [40]. This suggests that government agencies mainly adopt a top-down approach in risk communication and use the social media for one-way communication. The temporal changes of content patterns of personal posts indicate that the public have less interest in situation updates, general knowledge, policies, and guidelines as the epidemic evolves. However, the public may feel more empathic with the affected people and angry about other individuals or the government who put people at risk, as an increasing number of people are affected by the disease and the control measures. The government seemed to continue frequently sharing situation updates; policies, guidelines, and official actions; and prevention tips despite a decline in public interest. However, we also observed a significant increase in the frequency of instrumental and emotional support in government posts as the epidemic evolved. This provides in-depth understanding about why sentiment analysis indicates a decline in negative sentiment but increases in positive sentiment as the epidemic unfolded [35]. The inconsistent temporal changes in content patterns between personal and government posts and insufficient emotional support of government posts indicates an overall inadequate government responsiveness to public concerns. Reassuring the public about the epidemic risk was one of the main emotional contents identified from the government posts and was particularly apparent in the first 2 weeks; even Weibo data indicated a generally low risk awareness and concern among the general public. This is consistent with the public response to the 2009 influenza A/H1N1 pandemic shown by Twitter data [37]. Our study indicates that this may be because the government overreassured the public at the early epidemic stage, which is against Sandman and Lanard's [52] guidelines that risk communication should lean toward the alarming side particularly when the situation is uncertain. However, reassurance should be provided as more people are infected to deal with excessive or prolonged fear. The increasing use of blame in personal posts as more people were infected coincides with Douglas' [53] idea that contemporary risk is highly politicized. As more people are affected, "whose fault?" becomes a primary question to seek the accountability of certain persons or organizations and make sense of the epidemic [27]. Our study found that the public blamed not only individuals who put others at greater risk during the epidemic but also government, particularly local government figures for their perceived failures in risk communication and control measures. This also aligns with Beck's works on global risk society, in which authorities are increasingly questioned and blamed for failing to protect individuals [54]. Current data reveals seldom use of conspiracy theories in the attribution of blame and blame as a way to spread rumors [55]. In contrast, the government agencies tend to praise people who had made contributions to the control of the epidemic. This is viewed as a blame-avoidance strategy, called heroization [55], to direct public outrage to the appreciation of another group of people. Working with heroes in outbreak communication may be critical to improving communication effectiveness.

Limitations

First, individual posts were sampled by randomly choosing equal numbers of posts per calendar day for content analysis rather than using the common probability-based random sampling, which would generate a large sample size due to the vast amount of social media data. However, there is currently no consensus about how to best sample social media data, and our sampling strategy was able to draw a random manageable sample using manual coding. Second, only posts of government agencies that were more actively engaged with the public in online communication about COVID-19 were included for content analysis. This means that the sample of government posts for content analysis may not be representative of all government posts. However, this sample was considered to have greater impact on online audience's knowledge, perceptions, attitudes, and behaviors due to greater audience interest and attention to their messages. Third, our study did not evaluate the responses to and concerns over COVID-19 among nonnetizens, particularly those living in rural areas and older people in China [24], and the Chinese government's communications about COVID-19 through other channels. In addition, our data only covered the first 5 weeks after unknown pneumonia cases in Wuhan were first officially announced, which is a relatively short but critical period for outbreak communication. Furthermore, due to the problem of censorship on Weibo data in China, our data based on keyword extraction may lose a considerate part of Weibo data, particularly those posted before the official announcement of the outbreak in

Wuhan. This means that our data may not be able to accurately assess the delay of government responsiveness to the threat online.

Implications

The governments can closely monitor the social media discussions to identify public concerns to further improve government responsiveness in outbreak communication. To improve government responsiveness and public engagement, first, persons who have received training in risk communication can be designated for monitoring public concerns online, delivering timely messages, communicating about the uncertainty, and making efforts to address public concerns online. Second, the municipal health commission can communicate more locally relevant information to attract local followers' interest and motivate their information sharing. Third, the provincial and national health commissions can organize direct dialogues with online audiences on social media (eg, Weibo Chats) with trustworthy health care workers to capture audience's attentions and interests, and facilitate the rapid spread of fact-related messages [39]. Although the main role of the NHC in risk communication may be to disseminate facts, increasing messages showing empathy and care to affected people as the epidemic evolves is believed to be essential for maintaining credibility and trust in the public during a crisis [14].

Conclusion

The public seemed to respond earlier to the outbreak of COVID-19 online than government agencies. TThe Chinese government agencies' use of social media for outbreak communications remained limited to providing knowledge and information to the public. As the epidemic evolved, the public had declining interest in fact-related messages but became more empathic with the affected people and tended to attribute blame to other individuals or the government. The tendency of increasingly attributing blame to other individuals or the government may push the Chinese government to seek accountability and refine the compensation system for affected people. As more people are affected, the government may adopt a more empathic communication style to address public emotional response.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Descriptions of thematic categories of post contents.
[DOCX File , 16 KB - jmir_v22i5e18796_app1.docx]

Multimedia Appendix 2
Comparing model fit indices of latent class analysis models with different numbers of latent class by personal accounts and government agency accounts.
[DOCX File , 15 KB - jmir_v22i5e18796_app2.docx]

Multimedia Appendix 3
Item-response probabilities of each content pattern of the five-class model by personal accounts and government agency accounts.
[PNG File , 199 KB - jmir_v22i5e18796_app3.png]

References

1. Hong Kong Centre for Health Protection. 2019 Dec 31. CHP closely monitors cluster of pneumonia cases on Mainland URL: https://www.info.gov.hk/gia/general/201912/31/P2019123100667.htm
2. World Health Organization. 2020 Jan 09. WHO Statement regarding cluster of pneumonia cases in Wuhan, China URL: https://www.who.int/china/news/detail/09-01-2020-who-statement-regarding-cluster-of-pneumonia-cases-in-wuhan-china
3. World Health Organization. 2020 Feb 11. WHO Director-General's remarks at the media briefing on 2019-nCoV on 11 February 2020 URL: https://www.who.int/dg/speeches/detail/who-director-general-s-remarks-at-the-media-briefing-on-2019-ncov-on-11-february-2020
4. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb 15;395(10223):497-506. [doi: 10.1016/S0140-6736(20)30183-5] [Medline: 31986264]
5. Chan JF, Yuan S, Kok K, To KK, Chu H, Yang J, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet 2020 Feb 15;395(10223):514-523. [doi: 10.1016/S0140-6736(20)30154-9] [Medline: 31986261]
6. Wu P, Hao X, Lau EHY, Wong JY, Leung KSM, Wu JT, et al. Real-time tentative assessment of the epidemiological characteristics of novel coronavirus infections in Wuhan, China, as at 22 January 2020. Euro Surveill 2020 Jan;25(3):e. [doi: 10.2807/1560-7917.ES.2020.25.3.2000044] [Medline: 31992388]

7. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med 2020 Mar 26;382(13):1199-1207 [FREE Full text] [doi: 10.1056/NEJMoa2001316] [Medline: 31995857]
8. World Health Organization. 2020 May 13. Coronavirus disease (COVID-19) situation report – 114 URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200513-covid-19-sitrep-114.pdf?sfvrsn=17ebbbe_4 [accessed 2020-05-14]
9. Brooks SK, Dunn R, Amlôt R, Rubin GJ, Greenberg N. A systematic, thematic review of social and occupational factors associated with psychological outcomes in healthcare employees during an infectious disease outbreak. J Occup Environ Med 2018 Mar;60(3):248-257. [doi: 10.1097/JOM.0000000000001235] [Medline: 29252922]
10. Qiu J, Shen B, Zhao M, Wang Z, Xie B, Xu Y. A nationwide survey of psychological distress among Chinese people in the COVID-19 epidemic: implications and policy recommendations. Gen Psychiatr 2020;33(2):e100213 [FREE Full text] [doi: 10.1136/gpsych-2020-100213] [Medline: 32215365]
11. Liao JQ, Fielding R. Uncertain news: trust and preventive practices in respiratory infectious diseases. Eur Psychol 2014 Jan 01;19(1):4-12. [doi: 10.1027/1016-9040/a000168]
12. Depoux A, Martin S, Karafillakis E, Preet R, Wilder-Smith A, Larson H. The pandemic of social media panic travels faster than the COVID-19 outbreak. J Travel Med 2020 May 18;27(3):e [FREE Full text] [doi: 10.1093/jtm/taaa031] [Medline: 32125413]
13. Gentili D, Bardin A, Ros E, Piovesan C, Ramigni M, Dalmanzio M, et al. Impact of communication measures implemented during a school tuberculosis outbreak on risk perception among parents and school staff, Italy, 2019. Int J Environ Res Public Health 2020 Feb 01;17(3):e. [doi: 10.3390/ijerph17030911] [Medline: 32024183]
14. Reynolds B, Quinn Crouse S. Effective communication during an influenza pandemic: the value of using a crisis and emergency risk communication framework. Health Promot Pract 2008 Oct;9(4 Suppl):e. [doi: 10.1177/1524839908325267] [Medline: 18936255]
15. World Health Organization. 2005. WHO outbreak communication guidelines URL: https://apps.who.int/iris/bitstream/handle/10665/69369/WHO_CDS_2005_28_eng.pdf;jsessionid=3AB162F82325480AB56AE943A01F8E49?sequence=1 [accessed 2020-03-11]
16. Dickmann P, Abraham T, Sarkar S, Wysocki P, Cecconi S, Apfel F, et al. Risk communication as a core public health competence in infectious disease management: development of the ECDC training curriculum and programme. Euro Surveill 2016;21(14):e. [doi: 10.2807/1560-7917.ES.2016.21.14.30188] [Medline: 27103616]
17. Figueroa ME. A theory-based socioecological model of communication and behavior for the containment of the Ebola epidemic in liberia. J Health Commun 2017;22(sup1):5-9. [doi: 10.1080/10810730.2016.1231725] [Medline: 28854136]
18. Walker BF, Adukwu EC. The 2013-2016 Ebola epidemic: evaluating communication strategies between two affected countries in West Africa. Eur J Public Health 2020 Feb 01;30(1):118-124. [doi: 10.1093/eurpub/ckz104] [Medline: 31177274]
19. Lohiniva A, Sane J, Sibenberg K, Puumalainen T, Salminen M. Understanding coronavirus disease (COVID-19) risk perceptions among the public to enhance risk communication efforts: a practical approach for outbreaks, Finland, February 2020. Euro Surveill 2020 Apr;25(13):e. [doi: 10.2807/1560-7917.ES.2020.25.13.2000317] [Medline: 32265008]
20. Zhang X, Wen D, Liang J, Lei J. How the public uses social media wechat to obtain health information in china: a survey study. BMC Med Inform Decis Mak 2017 Jul 05;17(Suppl 2):66. [doi: 10.1186/s12911-017-0470-0] [Medline: 28699549]
21. Mackert M, Mabry-Flynn A, Champlin S, Donovan EE, Pounders K. Health literacy and health information technology adoption: the potential for a new digital divide. J Med Internet Res 2016 Oct 04;18(10):e264. [doi: 10.2196/jmir.6349] [Medline: 27702738]
22. Song H, Omori K, Kim J, Tenzek KE, Morey Hawkins J, Lin W, et al. Trusting social media as a source of health information: online surveys comparing the United States, Korea, and Hong Kong. J Med Internet Res 2016 Mar 14;18(3):e25. [doi: 10.2196/jmir.4193] [Medline: 26976273]
23. Beck F, Richard J, Nguyen-Thanh V, Montagni I, Parizot I, Renahy E. Use of the internet as a health information resource among French young adults: results from a nationally representative survey. J Med Internet Res 2014 May 13;16(5):e128. [doi: 10.2196/jmir.2934] [Medline: 24824164]
24. China Internet Network Information Center. 2019 Aug. Statistical report on internet development in China URL: https://cnnic.com.cn/IDR/ReportDownloads/201911/P020191112539794960687.pdf [accessed 2020-02-14]
25. Liu X, Lu J, Wang H. When health information meets social media: exploring virality on Sina Weibo. Health Commun 2017 Oct;32(10):1252-1260. [doi: 10.1080/10410236.2016.1217454] [Medline: 27668831]
26. Liao C, You S, Cheng Y. Network information analysis reveals risk perception transmission in a behaviour-influenza dynamics system. Epidemiol Infect 2015 Jan;143(1):23-36. [doi: 10.1017/S0950268814000430] [Medline: 24650779]
27. Roy M, Moreau N, Rousseau C, Mercier A, Wilson A, Atlani-Duault L. Ebola and localized blame on social media: analysis of Twitter and Facebook conversations during the 2014-2015 Ebola epidemic. Cult Med Psychiatry 2020 Mar;44(1):56-79 [FREE Full text] [doi: 10.1007/s11013-019-09635-8] [Medline: 31214902]
28. D Agostino M, Mejía F, Brooks I, Marti M, Novillo-Ortiz D, de Cosio G. Fear on the networks: analyzing the 2014 Ebola outbreak. Rev Panam Salud Publica 2017;41:e134 [FREE Full text] [doi: 10.26633/RPSP.2017.134] [Medline: 31384264]

29. Larson HJ. Blocking information on COVID-19 can fuel the spread of misinformation. Nature 2020 Apr;580(7803):306. [doi: 10.1038/d41586-020-00920-w] [Medline: 32231320]
30. Budhwani H, Sun R. Creating COVID-19 stigma by referencing the novel coronavirus as the "Chinese virus" on Twitter: quantitative analysis of social media data. J Med Internet Res 2020 May 06;22(5):e19301. [doi: 10.2196/19301] [Medline: 32343669]
31. Liao Q, Wu P, Wing Tak Lam W, Cowling BJ, Fielding R. Trajectories of public psycho-behavioural responses relating to influenza A(H7N9) over the winter of 2014-15 in Hong Kong. Psychol Health 2019 Feb;34(2):162-180. [doi: 10.1080/08870446.2018.1515436] [Medline: 30430862]
32. Xinhuanet. 2019 Aug 20. Weibo reports robust Q2 user growth URL: http://www.xinhuanet.com/english/2019-08/20/c_138323288.htm [accessed 2020-03-05]
33. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11. [doi: 10.2196/jmir.1157] [Medline: 19329408]
34. Fung IC, Fu K, Chan C, Chan BSB, Cheung C, Abraham T, et al. Social media's initial reaction to information and misinformation on Ebola, August 2014: facts and rumors. Public Health Rep 2016;131(3):461-473 [FREE Full text] [doi: 10.1177/003335491613100312] [Medline: 27252566]
35. Zhao Y, Cheng S, Yu X, Xu H. Chinese public's attention to the COVID-19 epidemic on social media: observational descriptive study. J Med Internet Res 2020 May 04;22(5):e18825. [doi: 10.2196/18825] [Medline: 32314976]
36. Odlum M, Yoon S. What can we learn about the Ebola outbreak from tweets? Am J Infect Control 2015 Jun;43(6):563-571 [FREE Full text] [doi: 10.1016/j.ajic.2015.02.023] [Medline: 26042846]
37. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118. [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
38. Li J, Xu Q, Cuomo R, Purushothaman V, Mackey T. Data mining and content analysis of the Chinese social media platform Weibo during the early COVID-19 outbreak: retrospective observational infoveillance study. JMIR Public Health Surveill 2020 Apr 21;6(2):e18700. [doi: 10.2196/18700] [Medline: 32293582]
39. Dalrymple KE, Young R, Tully M. “Facts, not fear”: negotiating uncertainty on social media during the 2014 Ebola crisis. Sci Commun 2016 Jun 22;38(4):442-467. [doi: 10.1177/1075547016655546]
40. Ding H, Zhang J. Social media and participatory risk communication during the H1N1 flu epidemic: a comparative study of the United States and China. China Media Res 2010:a.
41. Lazard AJ, Scheinfeld E, Bernhardt JM, Wilcox GB, Suran M. Detecting themes of public concern: a text mining analysis of the Centers for Disease Control and Prevention's Ebola live Twitter chat. Am J Infect Control 2015 Oct 01;43(10):1109-1111. [doi: 10.1016/j.ajic.2015.05.025] [Medline: 26138998]
42. Hamad EO, Savundranayagam MY, Holmes JD, Kinsella EA, Johnson AM. Toward a mixed-methods research approach to content analysis in the digital age: the combined content-analysis model and its applications to health care Twitter feeds. J Med Internet Res 2016 Mar 08;18(3):e60. [doi: 10.2196/jmir.5391] [Medline: 26957477]
43. Zhou Z, Zhang H, Xie J. Data crawler for Sina Weibo based on Python. J Computer Applications 2014;11:3131-3134.
44. Haro-de-Rosario A, Sáez-Martín A, del Carmen Caba-Pérez M. Using social media to enhance citizen engagement with local government: Twitter or Facebook? N Media Soc 2016 May 09;20(1):29-49. [doi: 10.1177/1461444816645652]
45. Bonsón E, Royo S, Ratkai M. Facebook practices in Western European Municipalities. Adm Soc 2016 Jul 27;49(3):320-347. [doi: 10.1177/0095399714544945]
46. Kim H, Jang SM, Kim S, Wan A. Evaluating sampling methods for content analysis of Twitter data. Soc Media Soc 2018 May 02;4(2):205630511877283. [doi: 10.1177/2056305118772836]
47. Collins LM, Lanza ST. Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences. Hoboken, NJ: Wiley; 2010.
48. Li R, Xie R, Yang C, Frost M. Perceptions on the risk communication strategy during the 2013 avian influenza A/H7N9 outbreak in humans in China: a focus group study. Western Pac Surveill Response J 2016;7(3):21-28 [FREE Full text] [doi: 10.5365/WPSAR.2016.7.1.005] [Medline: 27757257]
49. The Lancet. COVID-19: fighting panic with information. Lancet 2020 Feb;395(10224):537. [doi: 10.1016/s0140-6736(20)30379-2]
50. Kim C, Yang S. Like, comment, and share on Facebook: how each behavior differs from the other. Public Relations Rev 2017 Jun 10;43(2):441-449. [doi: 10.1016/j.pubrev.2017.02.006]
51. Obamiro K, West S, Lee S. Like, comment, tag, share: Facebook interactions in health research. Int J Med Inform 2020 May;137:104097. [doi: 10.1016/j.ijmedinf.2020.104097] [Medline: 32092667]
52. Sandman PM, Lanard J. The Peter M. Sandman Risk Communication Website. 2004. Crisis communication I: how bad is it? How sure are you? URL: http://www.psandman.com/handouts/sand12a.pdf [accessed 2020-03-12]
53. Douglas M. Risk and Blame: Essays in Cultural Theory. London: Routledge; 1992.
54. Sørensen MP. Ulrich Beck: exploring and contesting risk. J Risk Res 2017 Aug;21(1):6-16. [doi: 10.1080/13669877.2017.1359204]

55. Atlani-Duault L, Ward JK, Roy M, Morin C, Wilson A. Tracking online heroisation and blame in epidemics. Lancet Public Health 2020 Mar;5(3):e137-e138. [doi: 10.1016/S2468-2667(20)30033-5] [Medline: 32085818]

Abbreviations

aBIC: adjusted Bayesian information criterion
AIC: Akaike information criterion
BIC: Bayesian information criterion
COVID-19: coronavirus disease
LCA: latent class analysis
MHC: Municipal Health Commission
NHC: National Health Commission
SARS: severe acute respiratory syndrome
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2

Edited by G Eysenbach; submitted 19.03.20; peer-reviewed by PJ Schulz, J Zeng; comments to author 05.04.20; revised version received 11.04.20; accepted 14.05.20; published 26.05.20.

Please cite as:
Liao Q, Yuan J, Dong M, Yang L, Fielding R, Lam WWT
Public Engagement and Government Responsiveness in the Communications About COVID-19 During the Early Epidemic Stage in China: Infodemiology Study on Social Media Data
J Med Internet Res 2020;22(5):e18796
URL: http://www.jmir.org/2020/5/e18796/
doi:10.2196/18796
PMID:32412414

Original Paper

Complex Reporting of the COVID-19 Epidemic in the Czech Republic: Use of an Interactive Web-Based App in Practice

Martin Komenda[1,2], PhD; Vojtěch Bulhart[1,2]; Matěj Karolyi[1,2], MSc; Jiří Jarkovský[1,2], PhD; Jan Mužík[1,2], PhD; Ondřej Májek[1,2], PhD; Lenka Šnajdrová[1,2], PhD; Petra Růžičková[1,2], MSc; Jarmila Rážová[3], PhD; Roman Prymula[3], Prof Dr; Barbora Macková[4], PhD; Pavel Březovský[4], PhD; Jan Marounek[3], MBA, MD; Vladimír Černý[5,6], Prof Dr; Ladislav Dušek[1,2], Prof Dr

[1]Institute of Biostatistics and Analyses, Faculty of Medicine, Masaryk University, Brno, Czech Republic
[2]Institute of Health Information and Statistics of the Czech Republic, Prague, Czech Republic
[3]Ministry of Health of the Czech Republic, Prague, Czech Republic
[4]National Institute of Public Health, Prague, Czech Republic
[5]Department of Anesthesiology, Perioperative Medicine and Intensive Care, Masaryk Hospital, Ústí nad Labem, Czech Republic
[6]Jan Evangelista Purkyne University, Ústí nad Labem, Czech Republic

Corresponding Author:
Martin Komenda, PhD
Institute of Biostatistics and Analyses
Faculty of Medicine
Masaryk University
Kamenice 3
Brno
Czech Republic
Phone: 420 725174940
Fax: 420 549492855
Email: komenda@iba.muni.cz

Abstract

Background: The beginning of the coronavirus disease (COVID-19) epidemic dates back to December 31, 2019, when the first cases were reported in the People's Republic of China. In the Czech Republic, the first three cases of infection with the novel coronavirus were confirmed on March 1, 2020. The joint effort of state authorities and researchers gave rise to a unique team, which combines methodical knowledge of real-world processes with the know-how needed for effective processing, analysis, and online visualization of data.

Objective: Due to an urgent need for a tool that presents important reports based on valid data sources, a team of government experts and researchers focused on the design and development of a web app intended to provide a regularly updated overview of COVID-19 epidemiology in the Czech Republic to the general population.

Methods: The cross-industry standard process for data mining model was chosen for the complex solution of analytical processing and visualization of data that provides validated information on the COVID-19 epidemic across the Czech Republic. Great emphasis was put on the understanding and a correct implementation of all six steps (business understanding, data understanding, data preparation, modelling, evaluation, and deployment) needed in the process, including the infrastructure of a nationwide information system; the methodological setting of communication channels between all involved stakeholders; and data collection, processing, analysis, validation, and visualization.

Results: The web-based overview of the current spread of COVID-19 in the Czech Republic has been developed as an online platform providing a set of outputs in the form of tables, graphs, and maps intended for the general public. On March 12, 2020, the first version of the web portal, containing fourteen overviews divided into five topical sections, was released. The web portal's primary objective is to publish a well-arranged visualization and clear explanation of basic information consisting of the overall numbers of performed tests, confirmed cases of COVID-19, COVID-19-related deaths, the daily and cumulative overviews of people with a positive COVID-19 case, performed tests, location and country of infection of people with a positive COVID-19 case, hospitalizations of patients with COVID-19, and distribution of personal protective equipment.

Conclusions: The online interactive overview of the current spread of COVID-19 in the Czech Republic was launched on March 11, 2020, and has immediately become the primary communication channel employed by the health care sector to present the current situation regarding the COVID-19 epidemic. This complex reporting of the COVID-19 epidemic in the Czech Republic also shows an effective way to interconnect knowledge held by various specialists, such as regional and national methodology experts (who report positive cases of the disease on a daily basis), with knowledge held by developers of central registries, analysts, developers of web apps, and leaders in the health care sector.

(J Med Internet Res 2020;22(5):e19367) doi:10.2196/19367

KEYWORDS

coronavirus disease; COVID-19; Czech Republic; web app; interactive reporting; epidemiological overview; CRISP-DM; public health; app; epidemiology; virus; health data; data mining; modeling

Introduction

In early 2020, the pandemic of the coronavirus disease (COVID-19) started to spread all over the world. COVID-19 is caused by a novel type of coronavirus, referred to as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). This highly infectious disease is mainly manifested by fever, respiratory difficulty (cough, dyspnea), muscle pain, and fatigue. The disease can be rather serious for people who are older or have chronic illnesses and can even be fatal. The beginning of the COVID-19 epidemic dates back to December 31, 2019, when the first cases were reported in the city of Wuhan in Hubei Province in the People's Republic of China. Over the next 4 months, more than 1,120,000 people became infected across the world and almost 60,000 patients died from the disease [1]. In the Czech Republic, the first 3 cases of infection with the novel coronavirus were confirmed on March 1, 2020. A team of experts from the Institute of Health Information and Statistics of the Czech Republic (IHIS CR), together with researchers from the Institute of Biostatistics and Analyses at the Faculty of Medicine of the Masaryk University, focused on the design and development of a web app intended to provide a regularly updated overview of COVID-19 epidemiology in the Czech Republic to the general public. The joint effort of state authorities and researchers gave rise to a unique team, which combines methodical knowledge of real-world processes with the know-how needed for effective processing, analysis, and online visualization of data. The requirements on functionalities of this tool were mainly defined by the seriousness of an unexpected situation; an urgent need emerged for a tool that would make it possible to present important reports based on valid data sources only. To achieve this goal, it was necessary to ensure that individual graphs, maps, and tables would be easy to understand and could be correctly interpreted by the public and the media, and that misinterpretation of outputs would be avoided. An extensive review of tools available abroad was a valuable input for the development of a Czech tool. Worldwide, most papers published until recently [2-10] have been focused on the research of populations affected by COVID-19, on the structure of SARS-CoV-2 and its comparison with similar viruses (Middle East respiratory syndrome-related coronavirus and severe acute respiratory syndrome-related coronavirus), and on treatment and the overall mortality of COVID-19. From the technological point of view, several articles have been dedicated to the issue of data collection and sharing, together with their online presentation [1,11-15]. There are several examples of portals documenting the COVID-19 epidemiology clearly on a nationwide level, such as the Icelandic COVID-19 in Iceland – Statistics; the Korean Coronavirus Disease-19, Republic of Korea; or the Singaporean Dashboard of the COVID-19 Virus Outbreak in Singapore.

The aim of the research team was to adopt one of the time-tested methodologies for data mining, analytics, knowledge discovery, and data science projects, and to apply it in the process of mapping the current COVID-19 epidemic situation in the Czech Republic. This paper describes all essential steps from methodological as well as technical points of view. The Czech approach to the design, development, and implementation of online monitoring of the COVID-19 epidemic is based on a verified methodology for the acquisition, processing, and presentation of information. The methodology discussed in this paper made it possible to interconnect knowledge held by various specialists such as regional and national methodology experts from the National Institute of Public Health and regional public health authorities, who report positive cases of the disease on a daily basis, with knowledge held by developers of central registries, analysts, developers of web apps, and leaders in the health care sector.

Methods

Methodological Background for the COVID-19 App: An Overview of the Current Situation in the Czech Republic

When the urgent need emerged to map the current situation regarding the COVID-19 epidemic, it was essential to consider several factors that might have a significant impact on the resulting reports. The infrastructure of a nationwide information system run by the public administration is one of the most important factors in this regard; data from basic registries as well as data from health service providers are collected and processed in this information system. The cooperation between organizations involved in the process of data reporting, collection, processing, validation, analysis, evaluation, and visualization is another key aspect. The methodological setting of communication channels, duties resulting from legal measures, clearly defined competence, the sequence of steps to be made, and the overall management of the health care sector are complex at the time of an unexpected pandemic. In addition, it is important to select appropriate procedures for knowledge mining from database structures and for an undistorted

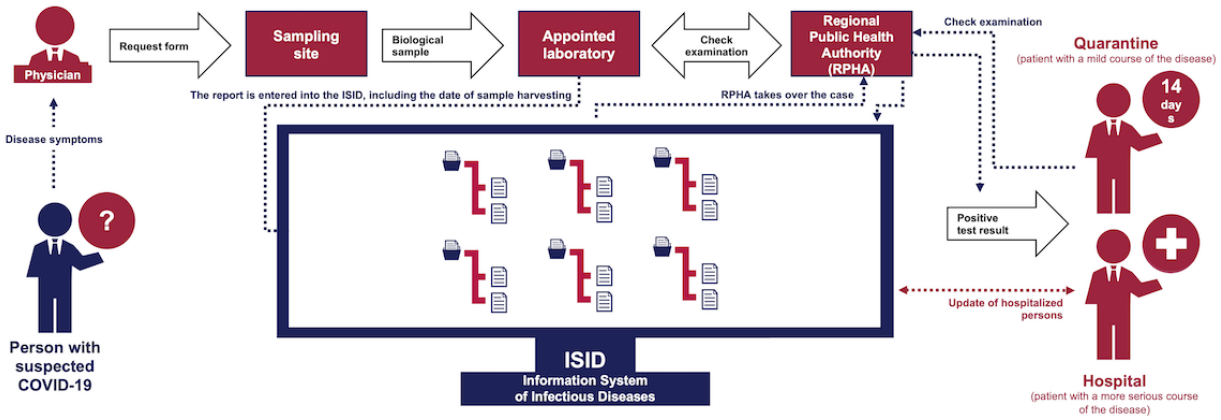
interpretation of data provided to individual target groups. These groups involve not only the general public and health care professionals but also the media, as the resulting reports are published on behalf of the Ministry of Health of the Czech Republic and can, therefore, be considered as guaranteed and entirely reliable.

Among several known methodological recommendations and standardized procedures, which can be cited as the implementation of the knowledge discovery in databases, the process Sample, Explore, Modify, Model, Assess and cross-industry standard process for data mining (CRISP-DM) are used most frequently in practice [16]. The team of authors chose the latter methodology (CRISP-DM) due to its higher versatility. The individual stages of this methodology directly correspond to the complex solution of analytical processing and visualization of data, which provides validated information on the COVID-19 epidemic across the Czech Republic. The power of CRISP-DM is demonstrated by the fact that great emphasis is put on the understanding and on a correct implementation of all six steps needed in the process (see Multimedia Appendix 1); furthermore, CRISP-DM provides the option to return back to previous steps, and importantly, it does not leave out the frequently omitted process of checking the achieved results before their publication.

Business Understanding

In the initial stage, maximum attention is paid to mapping the situation from the managerial point of view. In this case, the task is focused on online visualization of data on the current state of the COVID-19 epidemic in the Czech Republic. The Ministry of Health of the Czech Republic is responsible for the methodological setting of regular reports and processing data on newly identified cases of COVID-19 across the Czech Republic (see Figure 1), record keeping on testing locations, and overviewing the purchases and distribution of personal protective equipment (PPE). The reporting process always starts with a person with suspected COVID-19 being referred by a physician or a regional public health authority (RPHA) to a testing location, where a biological sample is taken from them, in compliance with the RPHA methodology. The sample is then analyzed by one of the appointed laboratories, which determines whether or not it is positive for SARS-CoV-2. The result is subsequently entered into the central Information System of Infectious Diseases (ISID), which is then taken over by the respective RPHA. In the next step, the same RPHA carries out a second investigation to verify the result, informs the person about the result, and, if the sample was confirmed to be positive for SARS-CoV-2, provides information about the next steps to be taken. The systematically designed architecture of the National Health Information System (NHIS) made it possible for the result of each performed test to be processed by a laboratory and verified by the respective RPHA in the central ISID, both procedures being done with only a minimum delay. The primary objective of the ISID is to obtain information on the incidence of infectious diseases to assess the epidemiological situation across the Czech Republic, to monitor the population's health status, and to control the provision of health care. In compliance with section 70, paragraph 3 of the Act No 372/2011 Coll, on Health Services and Conditions of Their Provision (Act on Health Services), the administration of NHIS has been delegated to the IHIS CR. NHIS is a fully computerized system involving components that are enshrined in the legislation. Each person in the Czech Republic can be unequivocally identified based on their birth certificate number; based on this number, essential links can be found between the NHIS and other relevant databases run by the public administration, such as the National Register of Hospitalized Patients, the Registry of Inhabitants, the National Register of Health Services Providers, the National Register of Health Care Professionals, or the Death Records Database. Data from these registries provide a comprehensive—and, most of all, up-to-date—data basis for subsequent analytical processing. On top of that, the unequivocal identification of patients, which is identical across these registries, made it possible to obtain information on the infection rate among health care workers in real time, for example.

Figure 1. Simplified schema of newly identified cases of COVID-19. COVID-19: coronavirus disease.

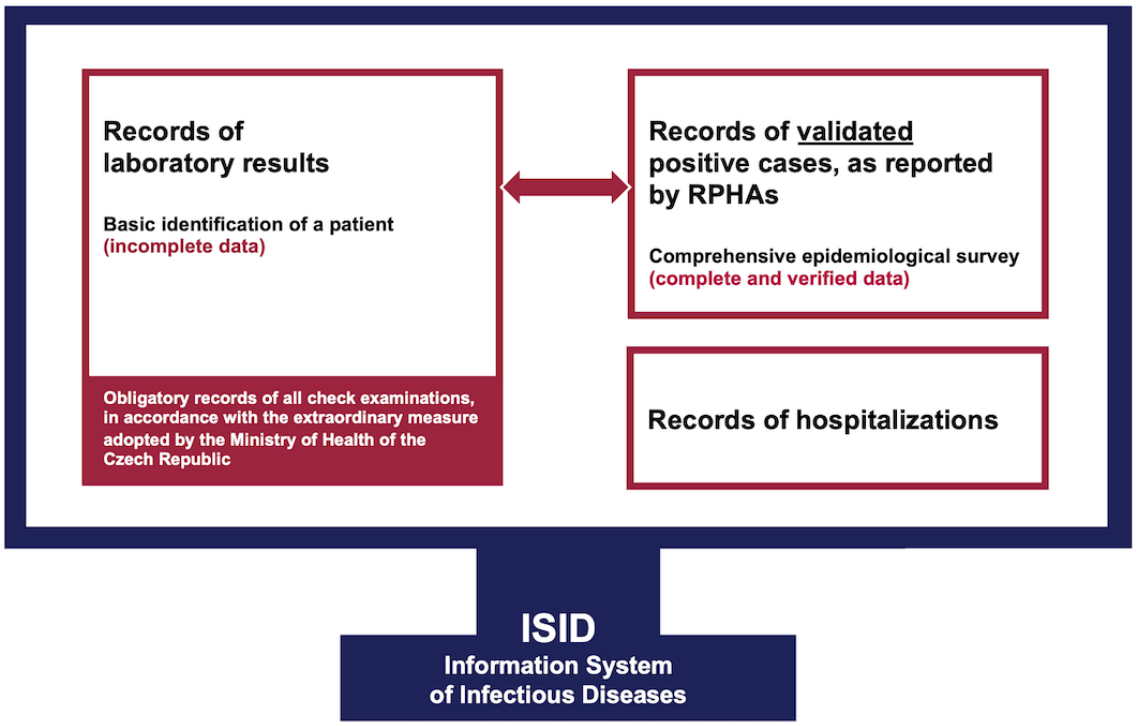


Data Understanding

For a correct collection and processing of information on the COVID-19 epidemiology in the Czech Republic, it is essential that all reporting processes are well understood, that communication with all involved institutions runs smoothly, that the architecture of central databases is well designed from the technical point of view, and that the resulting reports are correctly interpreted. At the same time, all interactive outputs,

whether in the form of graphs, maps, or tables, which are available for the general public, must meet the condition that it will never lead to a direct identification of any patient. It is therefore not possible to show detailed information on a district (or even a town) in combination with sex or age of a given person. What can be published, on the other hand, are summary data on the entire Czech Republic (or possibly on individual regions or districts), as the population of the entire country (or region or district) is large enough and the total number of positive cases is higher than 10, which cannot lead to the direct identification of a particular patient. Other examples of outputs that can be published involve daily reports on cumulative numbers and on the increase or decrease in the number of confirmed cases or division according to sex or age groups. When processing personal data in the various part of NHIS, every individual record must be processed in accordance with Regulation (EU) 2016/679 of the European Parliament and of the Council of April 27, 2016, on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), Act No 372/2011 Coll, on Health Services and Conditions of Their Provision (Act on Health Services), as subsequently amended, and the Act No 110/2019 Coll, on the Processing of Personal Data, as subsequently amended. The interactive online reporting is based on the five following data sets, which, combined together, characterize the COVID-19 epidemic in the Czech Republic. First, reports by RPHAs contain daily records of persons with confirmed COVID-19 (eg, date of report, age, sex, region, location, and country of infection). Second, reports by laboratories (LAB) contain deidentified records on persons with confirmed COVID-19, which have not yet been taken over by a RPHA in the respective region. In relation to reports by RPHAs, these are disjointed sets of records (each particular record goes to the LAB repository and consequently is either approved or directly confirmed by a respective RPHA); in other words, none of the records are present in both data sets, and undesirable duplicates are, therefore, avoided. Figure 2 shows how records provided by RPHAs and LABs are transferred into ISID. Among necessary adjustments, patient identification against the Registry of Inhabitants is performed at this stage, aiming to determine the region of a citizen who has been confirmed as a positive case by the laboratory. Third, a report on performed COVID-19 tests contains the number of all samples tested by laboratories across the Czech Republic on individual days. Fourth, reports on persons hospitalized in health care facilities contain daily summaries of currently hospitalized persons, persons in a serious condition or receiving highly intensive care (eg, mechanical ventilation, extracorporeal membrane oxygenation), and hospitalized persons who have been cured or discharged to home quarantine. Fifth, a report by the Ministry of Health of the Czech Republic contains purchases and distribution of PPE, and provides up-to-date numbers of face masks, respirators, goggles, bottles with disinfectant, face shields, and other equipment.

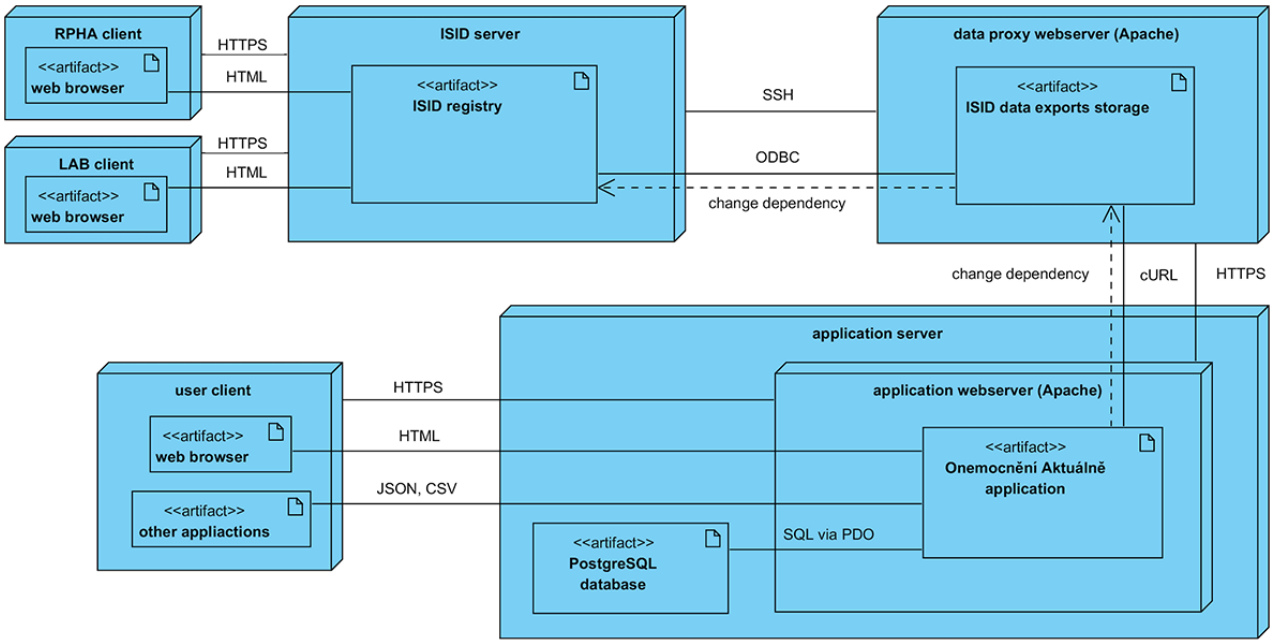
Figure 2. Diagram of data transfer within the Information System of Infectious Diseases.



Data Preparation

Several different data sources are used as input at this stage; these data sources must be validated thoroughly and then transformed into a form suitable for computer processing and subsequent visualization. In particular, the validation process involves a set of control mechanisms ensuring the completeness of individual records. This stage involves data cleaning (incomplete records are detected, corrected, or removed), construction of particular data views (selected attributes with primary identification are derived and merged), and data integration (final data sets are automatically generated). This process ensures that no invalid records are included in the stage of further processing and web visualization. Syncing tests, which are launched at regular intervals, are among the main control mechanisms. These tests provide information on whether or not data presented in the end report are consistent with data on the input. Their objective is to detect possible inconsistencies in the input records, such as a patient's incomplete or incorrectly entered birth certificate number; their permanent address; or distinction between a Czech citizen, a foreigner, or even a person who is homeless. If possible, standardized lists of values (ie, those valid on a nationwide level) are integrated, making it possible to name values in a unified way. As an example, a standardized classification of territorial units in the Czech Republic has been used (Nomenclature of Units for Territorial Statistics) based on unique codes for each of the 14 Czech regions (eg, CZ064 was used as a code for the South Moravian Region). Valid records in the ISID form the basis for the selection of descriptive attributes and their final adaptation into a format in which they are sent by export tools at regular intervals and in a secured manner to the web server. Figure 3 shows a diagram of implementation of individual components, showing the communication among three independent servers transmitting data sources that are necessary to draw the final visual outputs and to provide open data sets to the public.

Figure 3. Deployment diagram. CSV: comma-separated values; cURL: Client URL; ISID: Information System of Infectious Diseases; JSON: JavaScript Object Notation; LAB: laboratories; ODBC: Open Database Connectivity; PDO: PHP Data Objects; RPHA: regional public health authority; SQL; Structured Query Language; SSH: Secure Shell.



Modelling

In this step, detailed static instructions were prepared for each report (graph, map, table), always involving its data source, computational algorithm, description of visualization, explanation of its meaning, and information on the last update. Afterward, the entire process of data processing and dynamic drawing of the online report was transformed into a fully automatic one. All predefined data sets are transferred into an internal data storage device, which is dedicated to epidemiological analyses and reporting; individual overviews are subsequently generated in the form of interactive graphs, maps, and data tables. The resulting presentation is a set of interactive graph visualizations and tabular outputs, which can be viewed online by anyone using only a standard web browser. The portal has been developed using the Symfony 4.4 PHP framework. Graphical outputs are processed by the NVD3 JavaScript library, which is based on d3.js. Graphs are slightly adjusted to meet the report's needs, mainly in the responsive environment of the contemporary imaging technology. Tabular data are subsequently processed using the DataTables library, which internally employs jQuery (The jQuery Team). In this case, the data layer mainly consists of preprocessed and saved files in the JavaScript Object Notation (JSON) format. This approach significantly accelerates the access to data, making it possible to accelerate the app's response markedly. On each update, data are transformed straight away, source files being replaced by new ones. The number of computational operations that are carried out at the app layer is kept at a minimum. The presentation layer deals with the graphical appearance of the user interface with control elements as well as the interactivity of the visualization. All reports prepared in this way are first implemented on development servers, where functionality and data correctness are thoroughly verified. Development servers

are also used to check syncing with external data sources. Individual versions are subsequently published, always within the logical structure of the web app, which is divided into thematic sections. The app itself has been designed to be responsive and is fully supported by all types of devices, including the portable ones (mobile, tablet, and desktop).

Evaluation of Results

It is extremely important to validate the results before their publication, particularly when providing information about the spread of an epidemic. Despite several optimization and transformation processes performed in the preceding stages, data must remain consistent and all reports must correspond with the original input as well as with calculations carried out to check agreement with primary data. User testing not only revealed some inconsistency in original values obtained by static calculations, but also brought new ideas on how to improve the intelligibility, technical implementation, and user interface. Experts on accessibility of web apps were called in too. The target group was also taken into consideration, as it consists of disabled users, among others, for whom navigation on the internet might be difficult. A correct structure of headings, an adequate contrast between texts and the background, the overall legibility, a properly defined alternative information for images, and the availability of complementary tabular overviews next to graphs and maps are the most important accessibility attributes that have been thoroughly applied. One must keep in mind that web apps guaranteed by the state administration are required to provide maximum accessibility for all citizens without exception.

Deployment of Results

Primary data are entered into the ISID by LAB and RPHAs. After their validation and the unification of data formats, two data exports are created, and these are periodically sent from the ISID to a location from where they can be finally used for the purposes of the web portal. The periodicity of updates is set to 2 hours, with a nighttime shutdown between 2 am and 6 am The cron tool launches the updating script on the side of the "Disease at the Moment" ("Onemocnění aktuálně" in Czech) app at predefined times, which results in the transformation of provided data into preprocessed data sets. The previously mentioned syncing tests are launched during this process as well. The entire process is launched on the production instance of the "Disease at the Moment" app, as well as on the development instance and the stage instance. The latter two instances are automated by Jenkins, which is a continuous integration tool. In case of any error in data assembly or in the update of the entire app, an email notification is sent to the development team, allowing it to react appropriately. At the time of writing this paper (April 2020), primary data are updated three times a day on the production instance (12:30 am, 8:30 am, and 5:30 pm), and secondary data are usually updated once a day.

Open Data for COVID-19

Science is built on data, namely their collection, analysis, publication, reanalysis, critique, and reuse. Barriers include inability to access data, restrictions on usage applied by publishers or data providers, and publication of data that are difficult to reuse, for example, because they are poorly annotated or ''hidden'' in unmodifiable tables like PDF documents [17]. For that reason, the concepts of open access and open data are strongly emphasized and play a key role in the complex web-based reporting of the COVID-19 epidemic in the Czech Republic. The Ministry of the Interior of the Czech Republic guarantees and maintains the National Catalogue of Open Data, which consists of 24 local instances, including the Catalogue of Open Data run by the Ministry of Health of the Czech Republic. The main goal is to collect metadata about data sets published as open data throughout the whole country and to show transparency and effectiveness of government services [18]. There are three major target groups of potential users who can access various data sets freely: (1) those who are interested in information describing the entire concept together with benefits of broad usage of open data sets; (2) those who want to publish and update various data sets in accordance with the given open data rules, such as diagrams, data formats, keywords, and metadata description; and (3) those who can browse and freely use available data sets for the purpose of further analysis and development. The authors of this paper decided to adopt a standardized methodology for the publication of open data, and thus facilitate the use of COVID-19 epidemiological data sets by different public sector bodies, academic institutions, and business sector companies.

Google Analytics as the Monitoring Service

Monitoring and further analysis of web usage is one of the crucial points in terms of reflecting users' behavior and requirements. Systematic tracking and web analysis significantly improve the efficiency and quality during a long-term design and development of robust web apps. Google Analytics is a third-party service that measures and generates up-to-date statistics, reports, and analyses based on website traffic and on the behavior of its visitors. It tracks the visitors' activity, collects statistical data in real time, and stores them for a later analysis. Using different types of metrics, we can easily determine the number of visitors over any period, which pages they viewed, and how long was their visit. The flow of visitors is an important factor, showing user transitions between pages and the rate of abandonment of every single page [19]. Based on Google Analytics' powerful features, advanced visualizations mapping the analytics intelligence, dashboard, mobile device tracking, referrers, and geographic tracking capabilities have been used [20].

Results

Description of the COVID-19 App: An Overview of the Current Situation in the Czech Republic

On March 11, 2020, the first version of the web portal was released [21]. It provides a set of outputs in the form of tables, graphs, and maps intended for the general public and the media. Its primary objective is to provide a well-arranged visualization and clear explanation of basic information included in the basic overview of COVID-19 epidemiology in the Czech Republic (see Multimedia Appendix 2). The overall numbers of performed tests, confirmed cases of COVID-19, persons who have

recovered from the disease, and COVID-19-related deaths are displayed first to the user. The rest of the webpage is divided into several sections with coherent topics, each of them providing a synoptic overview from a selected perspective. First, basic overviews according to reports by RPHAs and positive results from laboratories show the overall number of persons who were tested positive for COVID-19, the incidence of confirmed cases of COVID-19 by region per 100,000 people, absolute numbers of people with positive COVID-19 cases by region, the overall number of performed tests (including repeated tests in the same person) for COVID-19 across the entire Czech Republic, and an overview of COVID-19-related deaths by age group and by region. Second, daily overviews according to reports by RPHAs and positive results from laboratories show a daily overview of number of persons with newly confirmed COVID-19, a daily overview of the number of performed tests (including repeated tests in the same person), the overall (cumulative) number of persons with laboratory-confirmed COVID-19 and the daily percentage change, and the daily percentage of persons with confirmed COVID-19 in the number of performed tests on a given day. Third, the number of persons with laboratory-confirmed COVID-19 according to reports by RPHAs shows an overview of cumulative numbers of persons with laboratory-confirmed COVID-19 that have been verified by RPHAs (not the number of all persons with laboratory-confirmed COVID-19), the location and country of infection of people with a positive COVID-19 case, and the number of persons with laboratory-confirmed COVID-19 by sex and age group. Fourth, an overview of hospitalizations of patients with COVID-19 shows the current number of hospitalized persons, the number of persons in a serious condition or receiving highly intensive care, and the number of hospitalized persons who have been cured or discharged to home quarantine. Fifth, an overview of the distribution of PPE shows how much PPE has been purchased and distributed by the Ministry of Health of the Czech Republic across individual regions.

Because published data are based on several independent sources, it is obvious that updates must be performed at different times. Nevertheless, fixed time intervals have been set at which updated values are presented on the web. The latest numbers summing up the incidence of COVID-19 in the Czech Republic are updated 3 times a day: at 12:30 am, at 8:30 am, and at 8:30 pm. The overviews showing total numbers from the previous day are prepared each morning at 8:30 am, and the latest numbers of recovered persons and COVID-19-related deaths are published at 8:30 am and at 5:30 pm.

Types of Visualization

In the stages of design and development of this web app, particular emphasis was placed on the character of presented data, which might be viewed by users anywhere and anytime. That was one of the reasons why responsive web design was among the main requirements, together with the overall optimization for mobile devices. The majority of line graphs and bar charts, therefore, primarily display an overview from the last 14 days, which is always adjusted even to small screens of mobile phones without the user having to manipulate the graph in any way. An additional complex view of the entire period of epidemic follow-up means that users of smaller devices may need to manipulate the graph if they want to display all values (Figures 4 and 5). A tabular overview is the last option available, displaying all values in rows and columns of a table. Other types of charts are also available on the website, such as the pie chart, stacked bar chart, or grouped bar chart, as well as map visualization and standard tables. Moreover, all published data are available to download in an open data format.

Figure 4. Daily trend in the percentage of persons with confirmed coronavirus disease in the overall number of persons tested on a given day.

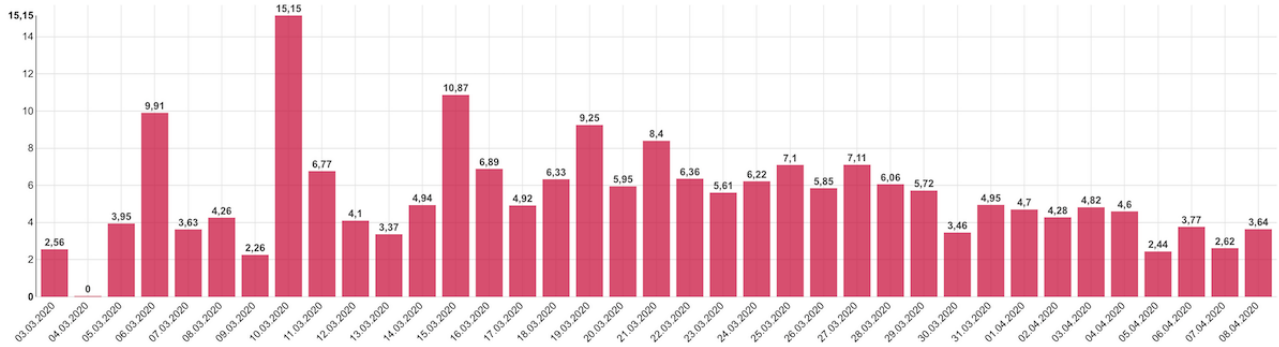
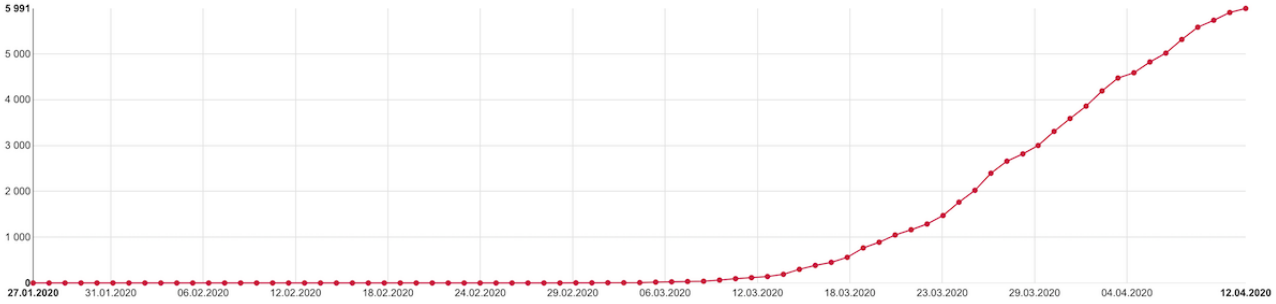


Figure 5. Overall (cumulative) number of persons with laboratory-confirmed coronavirus disease according to reports by regional public health authorities and laboratories.



Open Data Sets

Open data sets intended for further processing are an integral part of this information website. Source data in .CSV (comma-separated values) and .JSON formats are published daily at regular intervals, and can be downloaded by anyone who wants to process them either by a computer or manually. Data set schema are also included, describing the structure of records. Normalized schemas contain the list and definitions of all descriptive attributes as well as the technological representation of the data schema. All data sets published in this way are linked to records in the National Catalogue of Open Data (provided by the Ministry of Health of the Czech Republic), which is administered by the IHIS CR.

Analysis of the Number of Visits

The Google Analytics component is used to monitor and to analyze the users' behavior on the website. The objective of the development team was to monitor all basic characteristics such as the number of sessions, page views, ways of user acquisition (direct, organic, referral), type of device used (mobile, tablet, desktop), display resolution, and web browser. An overview of sessions on a given day and time makes it possible for the development team to plan timely releases of new versions of the portal and to update information. The website was officially launched on March 11, 2020, and has immediately become the primary communication channel employed by the health care sector to present the current situation regarding the COVID-19 epidemic. In the period between March 11, 2020, and April 12, 2020, the web portal registered 13,634,325 sessions and 19,662,351 page views. Returning visitors accounted for 16,286,551 (more than 80%) of users. This trend can also be confirmed by the graph showing how visitors landed on the website. A total of 7,158,020 (more than 50%) of visits were direct (ie, the website URL was typed into a browser) or were the result of an organic search (ie, users employed search engines such as Seznam.cz or Google.com). From the beginning, the development team designed the website as mobile-first. With regard to the portal contents, we assumed that its visitors would want to see the information fast and at any time. A total of 8,248,766 (more than 60%) of visitors used a mobile device or a tablet. A focus on graphs being drawn on displays with a certain resolution was no less important. The most usual width of mobile devices was 360 px, which comfortably accommodates a summary graph containing information from the last 14 days. Despite the clear preference of mobile devices, we also had to bear in mind that all graphs had to be well displayed in desktop browsers. Although Chrome was the most frequently used browser (8,112,423 of users, almost 60%), we also had to consider that some users prefer Internet Explorer (version 11.0 or lower), which accounted for more than 160,000 sessions. We used the Google Data Studio tool to create a dashboard that presents all needed information (see Multimedia Appendix 2) and thus provides a clear and easily available report. After the publication of open data sets and of a publicly available application programming interface on March 28, 2020, more than 100,000 page views were recorded over the next few days.

Discussion

The web-based app introducing an overview of the current spread of COVID-19 in the Czech Republic has been designed, developed, and implemented in accordance with the CRISP-DM methodology. All interactive graphs, maps, and tables fully respect strict rules of data management in the health care sector, where data reporting, collection, processing, validation, analysis, evaluation, and final publishing are under the supervision of the Ministry of Health of the Czech Republic. The online interactive overview of the current spread of COVID-19 in the Czech Republic [21] provides comprehensive information to the general public in a well-arranged manner. Since the launch of the first version of the website, the development team has not only systematically collected and evaluated suggestions on improvements from the general public but has also responded to the needs of leadership in the health care sector and of the media. Selected requirements (ie, those that are not contrary to legislation on personal data protection and that do not lead to a direct identification of an individual) are subsequently implemented and released in the next version. At the beginning of March 2020, the epidemiological situation in the Czech Republic was deemed rather serious, so many measures have been put in place aiming to curb the epidemic on a nationwide level and as effectively as possible. This is why an entire family of web apps focusing on COVID-19 has been under development. In addition to the existing overview of the current situation in the Czech Republic, two more online systems are planned to be launched in a short time (ie, in the next few weeks): (1) an epidemiological portal on COVID-19, which will present the descriptive demography, mathematical prediction models, overviews of incidence, prevalence and mortality, the R coefficient, and other relevant information; and (2) an online control room for intensive care, containing the latest reporting on occupancy and availability of beds in real

time, including an interface for a quick entry of the currently free capacity for patients with COVID-19 positive cases versus patients with COVID-19 negative cases. All three online tools have been primarily designed with the objective to keep everyone across the Czech Republic informed and to provide objective, data-based views for further decisions made by the leadership of the health care sector and by the emergency committee dealing with the COVID-19 epidemic in the Czech Republic.

Acknowledgments

This app was directly supported by the leadership of the Ministry of Health of the Czech Republic and the Institute of Health Information and Statistics of the Czech Republic. In particular, we would like to thank all regional public health authorities, laboratories, and hospitals for their enormous commitment during the COVID-19 epidemic.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Diagram of the cross-industry standard process for data mining reference model.
[PNG File , 166 KB - jmir_v22i5e19367_app1.png]

Multimedia Appendix 2
Selected visualization of coronavirus disease epidemiology in the Czech Republic. A: incidence of confirmed cases of coronavirus disease in the Czech Republic by region per 100,000 people. B: Overall number of persons who tested positive for coronavirus disease in the Czech Republic. C: Daily overview of the number of persons with newly confirmed coronavirus disease over the last 2 weeks. D: Overall number of persons with laboratory-confirmed coronavirus disease by age group.
[PNG File , 313 KB - jmir_v22i5e19367_app2.png]

Multimedia Appendix 3
Overview of monitored statistics regarding the number of visits (period: March 12-April 12, 2020).
[PNG File , 595 KB - jmir_v22i5e19367_app3.png]

References

1. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020 May;20(5):533-534 [FREE Full text] [doi: 10.1016/S1473-3099(20)30120-1] [Medline: 32087114]
2. Lipsitch M, Swerdlow DL, Finelli L. Defining the epidemiology of Covid-19 - studies needed. N Engl J Med 2020 Mar 26;382(13):1194-1196. [doi: 10.1056/NEJMp2002125] [Medline: 32074416]
3. Jiang F, Deng L, Zhang L, Cai Y, Cheung CW, Xia Z. Review of the clinical characteristics of coronavirus disease 2019 (COVID-19). J Gen Intern Med 2020 May;35(5):1545-1549. [doi: 10.1007/s11606-020-05762-w] [Medline: 32133578]
4. Jin H, Liu J, Cui M, Lu L. Novel coronavirus pneumonia emergency in Zhuhai: impact and challenges. J Hosp Infect 2020 Apr;104(4):452-453 [FREE Full text] [doi: 10.1016/j.jhin.2020.02.005] [Medline: 32068012]
5. Stebbing J, Phelan A, Griffin I, Tucker C, Oechsle O, Smith D, et al. COVID-19: combining antiviral and anti-inflammatory treatments. Lancet Infect Dis 2020 Apr;20(4):400-402 [FREE Full text] [doi: 10.1016/S1473-3099(20)30132-8] [Medline: 32113509]
6. Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. J Travel Med 2020 Mar 13;27(2):e [FREE Full text] [doi: 10.1093/jtm/taaa021] [Medline: 32052846]
7. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. 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] [Medline: 32109013]
8. Sun J, He W, Wang L, Lai A, Ji X, Zhai X, et al. COVID-19: epidemiology, evolution, and cross-disciplinary perspectives. Trends Mol Med 2020 May;26(5):483-495 [FREE Full text] [doi: 10.1016/j.molmed.2020.02.008] [Medline: 32359479]
9. Mavragani A. Tracking COVID-19 in Europe: infodemiology approach. JMIR Public Health Surveill 2020 Apr 20;6(2):e18941. [doi: 10.2196/18941] [Medline: 32250957]
10. Hernández-García I, Giménez-Júlvez T. Assessment of health information about COVID-19 prevention on the internet: infodemiological study. JMIR Public Health Surveill 2020 Apr 01;6(2):e18717. [doi: 10.2196/18717] [Medline: 32217507]
11. Moorthy V, Henao Restrepo AM, Preziosi M, Swaminathan S. Data sharing for novel coronavirus (COVID-19). Bull World Health Organ 2020 Mar 01;98(3):150 [FREE Full text] [doi: 10.2471/BLT.20.251561] [Medline: 32132744]
12. Heymann DL. Data sharing and outbreaks: best practice exemplified. Lancet 2020 Feb 15;395(10223):469-470. [doi: 10.1016/S0140-6736(20)30184-7] [Medline: 31986258]

13. Johns Hopkins University. 2020. OVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU) URL: https://coronavirus.jhu.edu/map.html [accessed 2020-04-10]
14. Worldometer. 2020. Coronavirus update (live) URL: https://www.worldometers.info/coronavirus/ [accessed 2020-04-10]
15. Our World in Data. 2020. Total confirmed COVID-19 cases URL: https://ourworldindata.org/grapher/total-cases-covid-19 [accessed 2020-04-10]
16. Azevedo A, Santos M. Repositório P.Porto. 2008. KDD, SEMMA and CRISP-DM: a parallel overview URL: http://recipp.ipp.pt/handle/10400.22/136 [accessed 2013-07-09]
17. Molloy JC. The Open Knowledge Foundation: open data means better science. PLoS Biol 2011 Dec;9(12):e1001195. [doi: 10.1371/journal.pbio.1001195] [Medline: 22162946]
18. Nečaský M, Klímek J, Škoda P. Practical use cases for linked open data in eGovernment demonstrated on the Czech Republic. In: Kő A, Francesconi E, editors. Electronic Government and the Information Systems Perspective. Switzerland: Springer, Cham; 2017:80-96.
19. Plaza B. Google Analytics for measuring website performance. Tourism Manage 2011 Jun;32(3):477-481. [doi: 10.1016/j.tourman.2010.03.015]
20. Gaur L, Singh G, Kumar S. Google Analytics: a tool to make websites more robust. 2016 Mar Presented at: Second International Conference on Information and Communication Technology for Competitive Strategies; 2016; Udaipur, India. [doi: 10.1145/2905055.2905251]
21. Komenda M, Karolyi M, Bulhart V, Žofka J, Brauner T, Hak J, et al. Onemocnění Aktuálně. COVID-19: overview of current situation in the Czech Republic URL: https://onemocneni-aktualne.mzcr.cz/covid-19 [accessed 2020-05-09]

Abbreviations

COVID-19: coronavirus disease
CRISP-DM: cross-industry standard process for data mining
CSV: comma-separated values
IHIS CR: Institute of Health Information and Statistics of the Czech Republic
ISID: Information System of Infectious Diseases
JSON: JavaScript Object Notation
LAB: laboratories
NHIS: National Health Information System
PPE: personal protective equipment
RPHA: regional public health authority
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2

Edited by G Eysenbach; submitted 15.04.20; peer-reviewed by F Dankar, M Pradhan, TH Kwan; comments to author 07.05.20; revised version received 14.05.20; accepted 14.05.20; published 27.05.20.

Please cite as:
Komenda M, Bulhart V, Karolyi M, Jarkovský J, Mužík J, Májek O, Šnajdrová L, Růžičková P, Rážová J, Prymula R, Macková B, Březovský P, Marounek J, Černý V, Dušek L
Complex Reporting of the COVID-19 Epidemic in the Czech Republic: Use of an Interactive Web-Based App in Practice
J Med Internet Res 2020;22(5):e19367
URL: http://www.jmir.org/2020/5/e19367/
doi:10.2196/19367
PMID:32412422

Viewpoint

Agile Health Care Analytics: Enabling Real-Time Disease Surveillance With a Computational Health Platform

Wade L Schulz[1,2], MD, PhD; Thomas J S Durant[1,2], MD; Charles J Torre Jr[1,3], MS; Allen L Hsiao[3,4], MD; Harlan M Krumholz[2,5,6], SM, MD

[1]Department of Laboratory Medicine, Yale School of Medicine, New Haven, CT, United States

[2]Center for Outcomes Research & Evaluation, Yale New Haven Hospital, New Haven, CT, United States

[3]Information Technology Services, Yale New Haven Health, New Haven, CT, United States

[4]Department of Pediatrics, Yale School of Medicine, New Haven, CT, United States

[5]Section of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, United States

[6]Department of Health Policy and Management, Yale School of Public Health, New Haven, CT, United States

Corresponding Author:
Wade L Schulz, MD, PhD
Department of Laboratory Medicine
Yale School of Medicine
55 Park St, PS215
New Haven, CT, 06511
United States
Phone: 1 (203) 688 2286
Email: wade.schulz@yale.edu

Abstract

The ongoing coronavirus disease outbreak demonstrates the need for novel applications of real-time data to produce timely information about incident cases. Using health information technology (HIT) and real-world data, we sought to produce an interface that could, in near real time, identify patients presenting with suspected respiratory tract infection and enable monitoring of test results related to specific pathogens, including severe acute respiratory syndrome coronavirus 2. This tool was built upon our computational health platform, which provides access to near real-time data from disparate HIT sources across our health system. This combination of technology allowed us to rapidly prototype, iterate, and deploy a platform to support a cohesive organizational response to a rapidly evolving outbreak. Platforms that allow for agile analytics are needed to keep pace with evolving needs within the health care system.

(J Med Internet Res 2020;22(5):e18707) doi:10.2196/18707

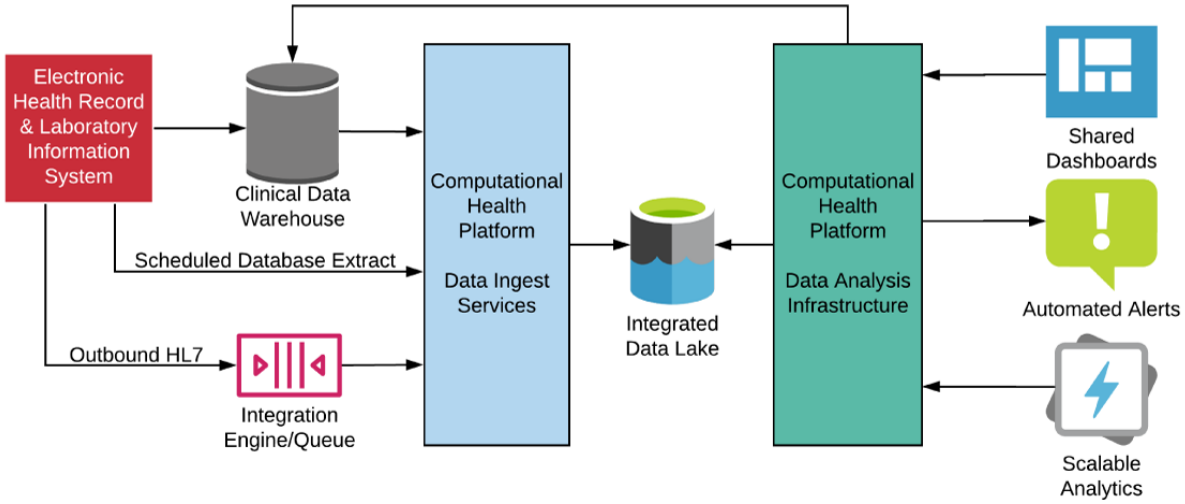
KEYWORDS

real-time analytics; real-world data; disease surveillance; computational health; surveillance; public health; COVID-19; outbreak; health information technology; HIT; interface; monitoring; pandemic

The ongoing coronavirus disease (COVID-19) outbreak demonstrates the need for novel applications of real-time data to produce timely information about incident cases [1]. Using health information technology (HIT) and real-world data (RWD), we sought to produce an interface that could in near real time, identify patients presenting with suspected respiratory tract infection and enable monitoring of test results related to specific pathogens, including severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).

This tool was built upon our computational health platform (CHP), which provides a data integration and analysis platform across our health system [2]. The platform enriches clinical data with geolocation information and real-time data from our enterprise integration engine. Data within our clinical data warehouses are transformed into the Observational Medical Outcomes Partnership (OMOP) common data model [3] and merged with data from the integration engine to support real-time analytics and promote the capacity for rapid collaboration. We developed an interactive dashboard to summarize these data and present them in a format suitable for analysis (Figure 1).

Figure 1. The architecture of data flows within the computational health platform with key integration points that allow for real-time data access and agile analytics. HL7: Health Level 7.



Factors impeding data access included limitations in electronic health record (EHR) interfaces and manual reporting workflows. After moving to an integrated EHR and laboratory information system, our Health Level 7 order and results feed became limited to results, which prevented access to other essential data such as order information, requiring us to develop a method to extract data from the underlying clinical database. The second bottleneck was the manual workflow in reporting COVID-19 results from the state reference laboratory, which initially relied on faxed paper results.

This real-time platform provides several benefits for managing a cohesive organizational response to a rapidly evolving outbreak. Trends can be tracked, issues with individual patients can be identified, hot spots can be determined, and new patients can be automatically reported for contact tracing. Delays in testing and resulting can also be identified across an integrated health care delivery system to address potential barriers to care delivery. Since these data are integrated with our OMOP data repository, downstream apps and timely observational research are also supported.

From March 1, 2020, through May 14, 2020, the tools described here were used to follow over 40,000 patients tested for COVID-19 across our health system, with information regarding admissions, intensive care unit capacity, and ventilator use for over 9600 patients with positive SARS-CoV-2 results. Dashboards were created to track laboratory testing volumes, turnaround times, and outstanding tests. The platform was also used to create a clinical COVID-19 data registry to support operational analytics, quality improvement initiatives, and biomedical research across the organization, and to create clinical predictive models [4].

With the rapid increase in COVID-19 and associated morbidity and mortality, there is a noted urgency to leverage available resources to identify risk factors and possible treatments, and track outcomes. For this, high-quality data are needed to generate evidence [5]. Although randomized controlled trials are being implemented internationally, they have several limitations, including typically long time periods from initiation to results. As such, other methods are needed to provide short-term information that can be used to guide the clinical and organizational response to a rapidly evolving pandemic. RWD has the potential to offer important insights into the management of COVID-19 and for the development of clinical predictive models that can be used in frontline care. Many investigators have begun to take advantage of RWD for COVID-19 research, but, although the information generated from RWD sources can be valuable, ensuring high-quality analysis requires new methodologic approaches, significant data validation, and careful interpretation of results.

Platforms that allow for agile analytics are needed to keep pace with evolving needs within the health care system. By leveraging our existing CHP, the initial tool was created and sourced with near real-time data in approximately 1 day of development effort from one developer and one clinical informaticist. These systems require an integrated team of HIT experts, clinical informaticists, health care providers, and administrative support to ensure that high-quality data can be efficiently disseminated across the enterprise. Such systems can be used to generate the RWD needed to drive evidence generation, especially with a rapidly evolving pandemic, but care is needed to ensure high-quality data are produced and that results are correctly interpreted given their observational nature.

Acknowledgments

We would like to acknowledge Nathan Price, William Byron, and Patrick Young for their assistance in developing the underlying platform and data feeds along with the ongoing updates to the analytic dashboard. We would like to acknowledge Phil Corso and Brian Keane for their assistance with ongoing analytic dashboard design and support.

Authors' Contributions

WS and TD conceived the study design, performed the analysis, and drafted and revised the manuscript. CT conceived the study design and revised the manuscript. AH and HK assisted with the study design and provided critical revision of the manuscript.

Conflicts of Interest

HK works under contract with the Centers for Medicare & Medicaid Services to support quality measurement programs; was a recipient of a research grant, through Yale University, from Medtronic and the US Food and Drug Administration to develop methods for postmarket surveillance of medical devices; was a recipient of a research grant with Medtronic and is the recipient of a research grant from Johnson & Johnson, through Yale University, to support clinical trial data sharing; was a recipient of a research agreement, through Yale University, from the Shenzhen Center for Health Information for work to advance intelligent disease prevention and health promotion; collaborates with the National Center for Cardiovascular Diseases in Beijing; receives payment from the Arnold & Porter Law Firm for work related to the Sanofi clopidogrel litigation, the Ben C Martin Law Firm for work related to the Cook Celect IVC filter litigation, and the Siegfried and Jensen Law Firm for work related to Vioxx litigation; chairs a Cardiac Scientific Advisory Board for UnitedHealth; was a participant/participant representative of the IBM Watson Health Life Sciences Board; is a member of the Advisory Board for Element Science, the Advisory Board for Facebook, and the Physician Advisory Board for Aetna; and is the cofounder of the personal health information platform HugoHealth and cofounder of Refactor Health, an AI-augmented data management platform for health care. WS was an investigator for a research agreement, through Yale University, from the Shenzhen Center for Health Information for work to advance intelligent disease prevention and health promotion; collaborates with the National Center for Cardiovascular Diseases in Beijing; is a technical consultant to HugoHealth, a personal health information platform, and cofounder of Refactor Health, an AI-augmented data management platform for health care; and is a consultant for Interpace Diagnostics Group, a molecular diagnostics company. The other coauthors report no potential conflicts of interest.

References

1. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020 May;20(5):533-534. [doi: 10.1016/s1473-3099(20)30120-1]
2. McPadden J, Durant TJ, Bunch DR, Coppi A, Price N, Rodgerson K, et al. Health care and precision medicine research: analysis of a scalable data science platform. J Med Internet Res 2019 Apr 09;21(4):e13043. [doi: 10.2196/13043] [Medline: 30964441]
3. Observational Health Data Sciences and Informatics. OMOP common data model URL: https://www.ohdsi.org/data-standardization/the-common-data-model/ [accessed 2019-03-02]
4. Haimovich A, Ravindra NG, Stoytchev S, Young HP, Wilson FP, van Dijk D, et al. Development and validation of the COVID-19 severity index (CSI): a prognostic tool for early respiratory decompensation. medRxiv 2020 May 14:a. [doi: 10.1101/2020.05.07.20094573]
5. Hope AA, Munro CL. Discussing evidence in the fog of the pandemic. Am J Crit Care 2020 May 05:e1-e3. [doi: 10.4037/ajcc2020239] [Medline: 32369547]

Abbreviations

CHP: computational health platform
COVID-19: coronavirus disease
EHR: electronic health record
HIT: health information technology
OMOP: Observational Medical Outcomes Partnership
RWD: real-world data
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2

Edited by G Eysenbach; submitted 12.03.20; peer-reviewed by T Ueno, A Benis; comments to author 20.04.20; revised version received 16.05.20; accepted 20.05.20; published 28.05.20.

Please cite as:
Schulz WL, Durant TJS, Torre Jr CJ, Hsiao AL, Krumholz HM
Agile Health Care Analytics: Enabling Real-Time Disease Surveillance With a Computational Health Platform
J Med Internet Res 2020;22(5):e18707
URL: http://www.jmir.org/2020/5/e18707/
doi:10.2196/18707
PMID:32442130

Original Paper

Using Reports of Symptoms and Diagnoses on Social Media to Predict COVID-19 Case Counts in Mainland China: Observational Infoveillance Study

Cuihua Shen[1*], PhD; Anfan Chen[2*], PhD; Chen Luo[3], MA; Jingwen Zhang[1,4], PhD; Bo Feng[1], PhD; Wang Liao[1], PhD

[1]Department of Communication, University of California, Davis, Davis, CA, United States

[2]Department of Science Communication and Science Policy, University of Science and Technology of China, Hefei, China

[3]School of Journalism and Communication, Tsinghua University, Beijing, China

[4]Department of Public Health Sciences, University of California, Davis, Davis, CA, United States

[*]these authors contributed equally

Corresponding Author:
Wang Liao, PhD
Department of Communication
University of California, Davis
One Shields Ave
Davis, CA,
United States
Phone: 1 5307520966
Email: wngliao@ucdavis.edu

Abstract

Background: Coronavirus disease (COVID-19) has affected more than 200 countries and territories worldwide. This disease poses an extraordinary challenge for public health systems because screening and surveillance capacity is often severely limited, especially during the beginning of the outbreak; this can fuel the outbreak, as many patients can unknowingly infect other people.

Objective: The aim of this study was to collect and analyze posts related to COVID-19 on Weibo, a popular Twitter-like social media site in China. To our knowledge, this infoveillance study employs the largest, most comprehensive, and most fine-grained social media data to date to predict COVID-19 case counts in mainland China.

Methods: We built a Weibo user pool of 250 million people, approximately half the entire monthly active Weibo user population. Using a comprehensive list of 167 keywords, we retrieved and analyzed around 15 million COVID-19–related posts from our user pool from November 1, 2019 to March 31, 2020. We developed a machine learning classifier to identify "sick posts," in which users report their own or other people's symptoms and diagnoses related to COVID-19. Using officially reported case counts as the outcome, we then estimated the Granger causality of sick posts and other COVID-19 posts on daily case counts. For a subset of geotagged posts (3.10% of all retrieved posts), we also ran separate predictive models for Hubei province, the epicenter of the initial outbreak, and the rest of mainland China.

Results: We found that reports of symptoms and diagnosis of COVID-19 significantly predicted daily case counts up to 14 days ahead of official statistics, whereas other COVID-19 posts did not have similar predictive power. For the subset of geotagged posts, we found that the predictive pattern held true for both Hubei province and the rest of mainland China regardless of the unequal distribution of health care resources and the outbreak timeline.

Conclusions: Public social media data can be usefully harnessed to predict infection cases and inform timely responses. Researchers and disease control agencies should pay close attention to the social media infosphere regarding COVID-19. In addition to monitoring overall search and posting activities, leveraging machine learning approaches and theoretical understanding of information sharing behaviors is a promising approach to identify true disease signals and improve the effectiveness of infoveillance.

(J Med Internet Res 2020;22(5):e19421) doi:10.2196/19421

KEYWORDS

COVID-19; SARS-CoV-2; novel coronavirus; infectious disease; social media; Weibo; China; disease surveillance; surveillance; infoveillance; infodemiology

Introduction

Since the outbreak of coronavirus disease (COVID-19) in December 2019 in Wuhan, Hubei Province, China [1,2], the novel coronavirus has affected more than 200 countries and territories worldwide. As of May 16, 2020, there were more than 4 million confirmed cases of COVID-19 and over 300,000 deaths [3]. Amid many unknown factors, severe lack of laboratory testing capacity, delays in case reports, variations in local COVID-19 responses, and uncoordinated communication pose tremendous challenges for monitoring the dynamics of the epidemic and developing policies and targeted interventions for resource allocation.

When conventional disease surveillance capacity is limited, publicly available social media and internet data can play a crucial role in uncovering the hidden dynamics of an emerging outbreak [4]. Research in digital disease surveillance, also referred to as infoveillance or infodemiology, has shown great promise in the useful employment of internet data to track the real time development of public attention, sentiment, and health [5-8]. Specifically, data based on internet searches and social media activities can nowcast and forecast disease prevalence as a supplement to conventional surveillance methods for various infectious diseases [5-7,9-14].

One of the best-known examples of digital disease surveillance is Google Flu Trends, which used real time Google search terms to predict clinical incidence rates of influenza with great initial success [13,14]. Data from social media platforms such as Twitter have also been shown to be effective in predicting and tracking various epidemics, such as influenza [10,12] and Zika virus [15], with varying degrees of success. However, digital surveillance data present unique challenges. For example, after its release in 2008, Google Flu Trends became less accurate over time, consistently overestimating flu prevalence during 2011-2013. The prediction error was partially attributed to people's changing search behaviors as well as increased public attention to the epidemic itself, which fueled awareness-related search queries that were not strongly related to disease incidence [7,16]. Compared to aggregated search queries, user-generated social media data have the advantage of being more direct and granular, allowing researchers to mine specific content to reflect actual illness. However, media attention to emerging diseases can fuel social media activities, resulting in a deluge of discussions that dilute true disease signals of actual infection cases; thus, predictions are less accurate [12].

The unprecedented magnitude and transmission speed of COVID-19 brought about massive social media activities as people isolated themselves in their homes to break the infection chain [17]. Massive social media data inevitably contain massive noise (eg, public reactions and awareness of the disease), which can be counterproductive for disease forecasting. A few early infoveillance studies tracked public discussion of COVID-19 and patient characteristics on Weibo, the most popular public social media site in China [18-21]. Two studies suggested that COVID-19–related Weibo posts and search queries can be used to predict disease prevalence [19,22]. However, these studies relied upon coarse-grained social media data and query data based on a few keywords with a short time window at the onset of the outbreak [19,22]. As such, the predictive accuracy and result interpretability of these studies are limited by the same pitfalls of infoveillance studies mentioned above. There are many reasons to search for and discuss COVID-19 on social media, especially because the disease has received substantial media coverage and many countries are under mandatory lockdown. To more accurately predict infection cases and inform a rapid response, it is therefore critical to use granular and specific social media data to identify reliable disease signals (ie, "sick posts" reporting symptoms and diagnosis).

Here, we present an infoveillance effort to collect and analyze COVID-19–related posts on Weibo and to identify specific types of Weibo posts that can predict COVID-19 case counts in mainland China. To our knowledge, this study involves the largest, most comprehensive, and most granular collection of social media data related to COVID-19 in the Chinese language, far exceeding the scale, granularity, and timespan of similar studies [19,22]. We built a Weibo user pool of 250 million people, approximately half the active Weibo user population [23]. Using a comprehensive list of 167 keywords associated with COVID-19, we retrieved around 15 million social media posts from November 1, 2019 to March 31, 2020. With greatly increased data granularity, we developed a supervised machine learning classifier to distinguish "sick posts," which are reports of one's own and other people's symptoms or diagnosis, from other COVID-19 related posts that could dilute disease signals from the data stream. Using the officially reported case counts as the outcome, we compared the predictive power of sick posts versus other COVID-19 posts. We show evidence that sick posts predicted the daily cases reported by the Chinese Center for Disease Control and Prevention (China CDC) up to 14 days in advance, while other COVID-19–related posts had much weaker predictive power. For the subset of geotagged posts, we found that the predictive pattern held true for both Hubei province and the rest of mainland China. Our work demonstrates a viable method to identify disease signals through reports of symptoms or diagnosis rather than relying upon general discussion of COVID-19, making a significant contribution to the infoveillance literature.

Methods

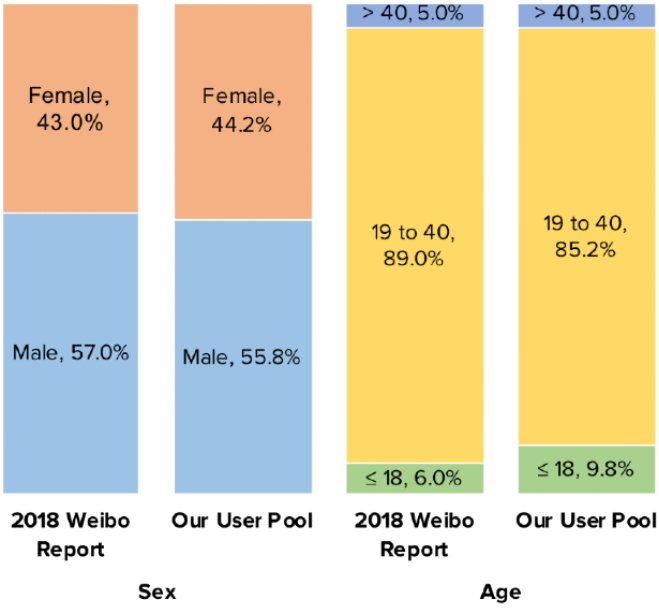
Data Collection

The social media data used in this study were collected from a popular Chinese microblog platform, Weibo, which had over 516 million monthly active users at the end of 2019 [23]. Weibo is very similar to Twitter, access to which is blocked in mainland China. Unlike Twitter, Weibo does not provide large-scale public application programming interface (API) access to its

database. Weibo enables keyword-based advanced searching of posts via its web interface; however, per Weibo policy, the output of these searches is limited to 50 pages (or around 1000 posts). Therefore, large-scale public data access is notoriously difficult.

To bypass these limitations, we employed a Weibo user pool originally built in 2018, which started from 5 million active Weibo users obtained in our previous research unrelated to COVID-19 [24,25]. We then retrieved the initial 5 million users' followers and followees (second degree users), the followers and followees of the second degree users (third degree users), etc., until no new users were found. This snowball process resulted in a pool of 250 million users (with bots filtered out), which represents approximately 48.4% of all monthly active Weibo users in 2019 [23] and is similar to the 2018 population of Weibo users in terms of self-reported sex and age distribution [26] (see Figure 1).

Figure 1. Demographic composition of our Weibo user pool compared to that in the 2018 Annual Sina Weibo user report. Age is reported in years.



COVID-19 Posts

Following best practices for content retrieval and analysis [27], we generated a comprehensive list of keywords related to COVID-19 through close observation of Weibo posts every day from late January to March 2020. We then retrieved COVID-19 posts by searching all posts by users in the user pool with 167 keywords covering general terms related to the epidemic, such as coronavirus and pneumonia, as well as specific locations (eg, "Wuhan"), drugs (eg, "remdesivir") and preventive measures (eg, "mask"). For a complete keyword list, see Multimedia Appendix 1, Table A).

After removing duplicates (ie, reposts of original posts), we retained 14,983,647 posts sent between November 1, 2019 (ie, 30 days before the first confirmed cases) and March 31, 2020 (to access the Weibo dataset on COVID-19, see [28]).

A subset of 464,111/14,983,647 of these posts (3.10%) were tagged with geographic information. We distinguished between posts sent within Hubei province (ie, the epicenter; 169,340/14,983,647; 36.49%) and those from elsewhere in mainland China (294,771/14,983,647; 63.51%).

Sick Posts

We conceptually defined "sick posts" as posts that report any symptoms or diagnoses that are likely related to COVID-19 based on published research and news reports from the medical social media site DXY.cn [29]. We collected a broad list of symptoms, including common symptoms such as cough and shortness of breath and uncommon symptoms such as diarrhea. Sick posts can be further categorized into "ingroup sick posts," which we defined as posts that disclose the user's own or immediate family members' symptoms or diagnoses, and "outgroup sick posts," which report symptoms and diagnoses of people not in the user's immediate family. The reason for the a priori categorization is that people tend to have firsthand and more accurate information about their own or immediate family members' medical conditions; meanwhile, they have much less reliable information about people outside of their household, especially during a national lockdown. All posts that were obtained using the 167 keywords but did not fall into these categories were classified as "other COVID-19 posts." We provide an example of an ingroup sick post below (translated and edited for brevity):

> *During the SARS epidemic in 2003, I got pneumonia with symptoms of fever and cough, was suspected of being infected with SARS, and ended up being hospitalized for more than a month. Now we got COVID-19 in 2020 and I started coughing again, which has lasted for more than a month. What a mess <Face Palm> (Posted 10:23 PM, January 29, 2020)*

We also provide an example of an outgroup sick post:

> *One man in another village drank too much. He said he felt sick and had cold symptoms. His brother measured his temperature which turned out to be 38 Celsius. His brother called 120 and sent him to*

hospital. The whole village was shocked and everyone was afraid to go outside. (Posted 10:14 PM, January 29, 2020)

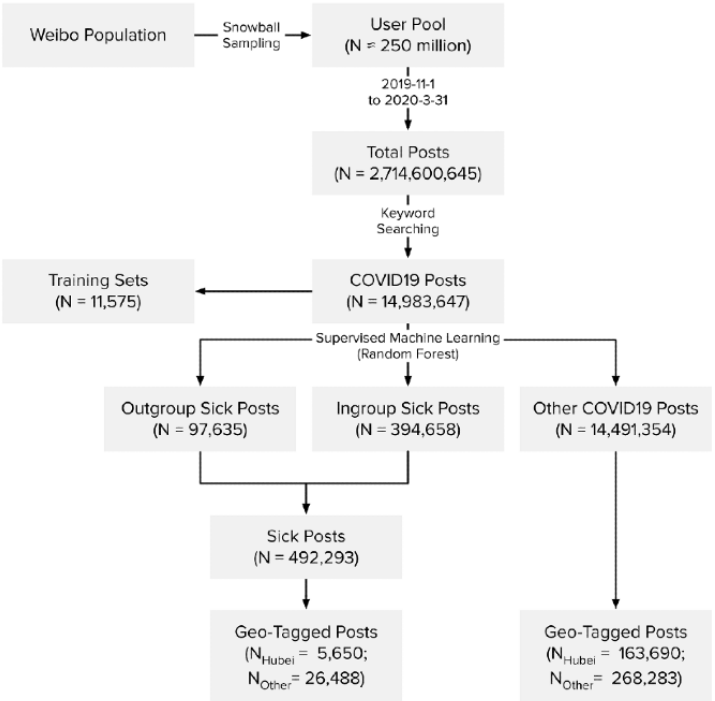
We used supervised machine learning algorithms to identify sick posts from the keyword-retrieved COVID-19 posts. We first sampled 11,575 posts in proportion to the retrieved posts across 5 months of data collection. Next, 11 human judges annotated whether a post was an ingroup sick post, outgroup sick post, or other COVID-19 post. The judges independently annotated a subset of 138 posts and achieved high agreement (Krippendorff α=.945) before they divided and annotated the remaining posts. Then, the annotated posts were used to train machine learning models with various algorithms. Based on the classification performance (see Table 1), we selected the model using the random forest algorithm (F1 score=0.880). The model classified the 14,983,647 COVID-19 posts into 394,658 (2.63%) ingroup sick posts, 97,635 (0.65%) outgroup sick posts, and 14,491,354 (96.71%) other COVID-19 posts. Because of the low number of outgroup sick posts, we combined ingroup and outgroup sick posts in subsequent analyses.

Among the subset of geotagged COVID-19 posts (464,111/14,983,647, 3.10% of all retrieved posts), 5,650 sick posts (1.2%) and 163,690 other COVID-19 posts (35.3%) were tagged in Hubei; meanwhile, 26,488 sick posts (5.7%) and 268,283 other COVID-19 posts (57.8%) were from elsewhere in mainland China. These post counts were then aggregated by days. To control for the day-to-day fluctuations of Weibo posts, we further normalized these numbers against the daily counts of all Weibo posts generated by our user pool. The normalized sick post and other COVID-19 post counts can be interpreted as counts per 1 million posts. Figure 2 summarizes our data collection and classification process.

Table 1. Performance of machine learning models in classifying sick posts.

Model	F1 score	Precision	Accuracy	Recall
Decision tree	0.835	0.840	0.830	0.830
Extra tree	0.785	0.785	0.785	0.785
Extra trees	0.878	0.881	0.885	0.885
K nearest neighbors	0.810	0.819	0.819	0.819
Multilayer perceptron	0.847	0.845	0.851	0.851
Support vector machine	0.877	0.877	0.878	0.878
Random forest	0.880	0.885	0.888	0.888

Figure 2. Weibo data collection and classification procedure.



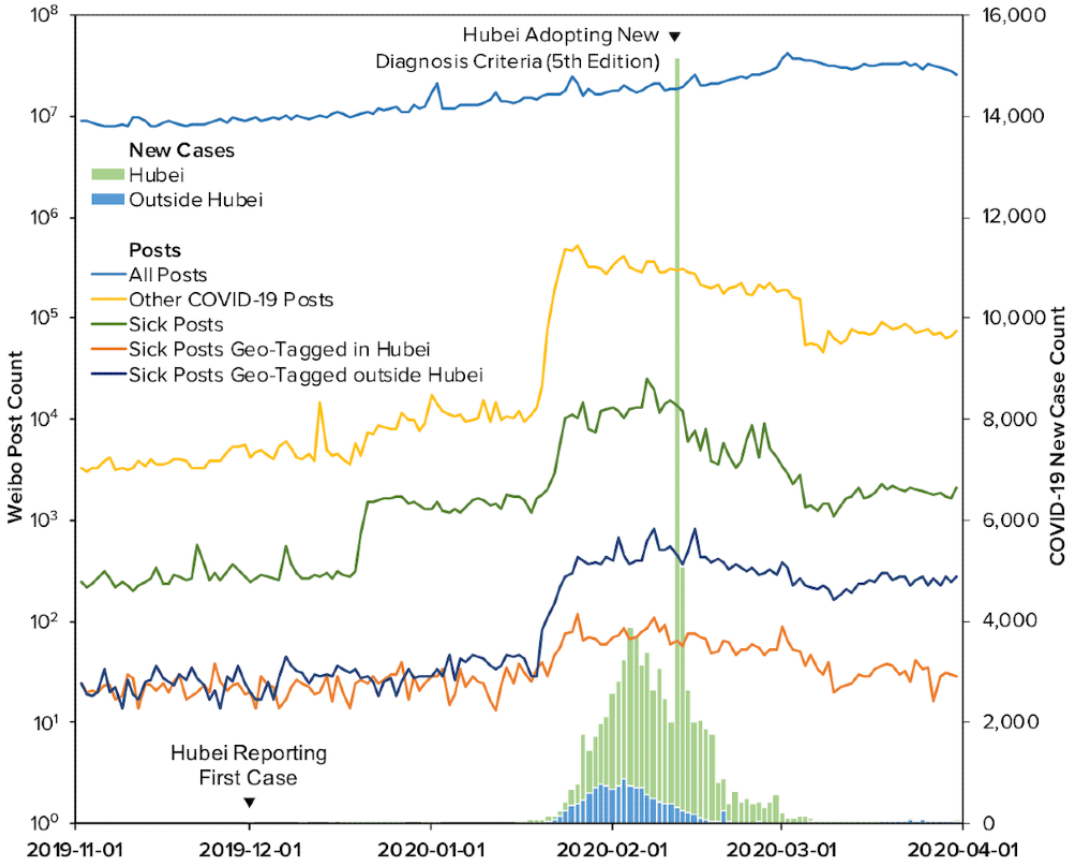
COVID-19 Daily Case Counts

We collected the daily new case counts in mainland China from China CDC on May 8, 2020. China CDC's official website started collating data on January 16, 2020. Earlier counts were obtained from Huang et al [1] and validated against relevant briefings from the National Health Commission. The final case data cover the same period from November 1, 2019, to March

31, 2020, within which the first reported COVID-19 clinical case dates back to December 1, 2019. We also distinguished between cases within and outside Hubei (see Figure 3).

It is noteworthy that China CDC released seven editions of diagnostic criteria throughout the time period covered in this study and thus introduced systematic changes to the case counts. Particularly, on February 12, 2020, Hubei province started to implement the fifth edition of the COVID-19 diagnostic criteria released on February 4, 2020. This led to a temporary surge of new cases [30]. The impact of this incident was controlled for in our analyses, as discussed in the section below. After close comparison of each edition, we concluded that the changes among other editions of the diagnostic criteria were relatively minor, and their release dates did not appear to be associated with abrupt changes in the case counts; therefore, we did not further control for them.

Figure 3. Daily Weibo posts and confirmed COVID-19 cases between November 1, 2019 and March 31, 2020.



Statistical Analysis

We performed Granger causality tests [31] to discover if an increase of sick posts forecasted an increase of new cases, as formulated in the following linear model:

where Ct is the difference in new case counts at day t from day $t-1$, S_{t-i} is the difference in sick post counts (normalized) at day t from day $t-1$, and I_t is a time-varying binary variable that equals 1 on February 12, 2020, the day on which Hubei adopted the fifth edition of the diagnostic criteria. This binary variable controls for the exogenous pulse of case counts [32]. Since we collected Weibo posts from as early as November 1, 2019, 30 days before the first reported case of COVID-19 on December 1, 2019, we were able to test up to 29 lags of such posts (ie, $m \leq 29$). The model is further explained as follows.

First, difference scores instead of raw new case counts were used because Dickey-Fuller tests for the raw counts could not reject nonstationarity (ie, the presence of a unit root) for lag 3–29 at a 5% confidence level (see Table B in Multimedia Appendix 1). Both stationarity and the inclusion of autoregressive terms are required by Granger causality. In contrast, the Dicky-Fuller tests suggested that the difference scores of the case counts were stationary: nonstationarity was rejected for lag 1–12 at a 1% confidence level and for lag 13–29 at a 5% confidence level (see Table B in Multimedia Appendix 1). The Dickey-Fuller tests reached the same conclusion for the stationarities of the sick post counts and their difference scores (see Table B in Multimedia Appendix 1). We thus also used the difference scores instead of the raw counts to reduce correlations among lag terms of sick post counts. This more clearly identifies their independent effects on case counts. In short, these difference scores can be interpreted as "daily-additional" cases or Weibo posts in addition to the counts from the previous day.

Second, to determine the number of lag terms to include (ie, m in the above formula), we compared model fit statistics while iteratively adding lag terms. The model comparison suggested that the inclusion of more lags continuously improved the model fit up to the maximum lags (ie, 29; see Table C in Multimedia Appendix 1). However, the parameter estimates did not change

qualitatively after including more than 20 lags (see Tables D and E in Multimedia Appendix 1). For parsimony and statistical power, we settled at 20 lags for the following analyses.

Finally, we included a binary variable to control for the change in the diagnostic criteria of COVID-19 on Feb 12, 2020, following the procedure of intervention analysis [33]. Because this change is unlikely to induce permanent changes to case counts, an instant pulse function was applied at the date of the change. We also tested models that allowed the effect to linearly decay in 2, 3, 4, or 5 days; these models fitted the data more poorly than the model with an instant pulse (see Table F in Multimedia Appendix 1).

Results

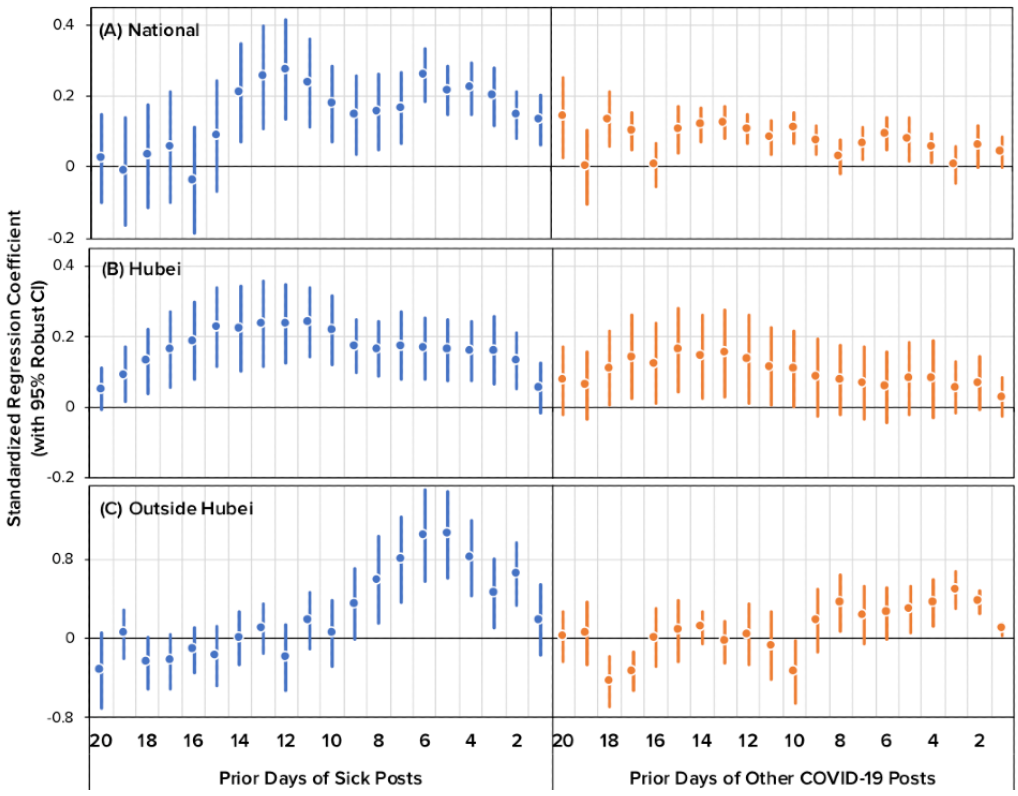
Ordinary least squares regression with robust standard errors was used to estimate the final models. With 20 lag terms in the model, the modeled data include daily-additional new COVID-19 cases from December 1, 2019 to March 31, 2020 and daily-additional counts of sick posts and other COVID-19 posts from November 10, 2019 to March 11, 2020 (N=122).

Figure 4A summarizes the estimates of Granger causality for sick posts predicting new COVID-19 cases with standardized regression coefficients (see Table G in Multimedia Appendix 1 for all estimated parameters). Particularly, one standard deviation of increase in the daily-additional sick posts (1 sick post per 1 million posts) predicted a 0.133 (95% CI 0.065-0.201) to 0.275 (95% CI 0.134-0.416) standard deviation of the increase in the daily-additional new cases 1-14 days in advance. After including the 20 lags of sick posts, the adjusted R^2 value of the model increased by 0.128, suggesting that sick posts could explain an additional 12.8% of the variance of daily-additional new cases beyond the autoregressive terms and intervention effects.

Furthermore, we estimated the relationship between other COVID-19 post counts and daily-additional new cases using the same linear model. Figure 4A further illustrates the standardized estimates. Compared with sick posts, other COVID-19 posts were weaker signals of future case counts, as demonstrated by their smaller standard regression coefficients. This indicates that Weibo posts that discussed some aspect of COVID-19 but did not explicitly report a person's symptoms or diagnosis had lower forecasting power than sick posts.

To corroborate the above results, we tested the Granger causality of sick posts on cases within Hubei and outside Hubei (see Table H in Multimedia Appendix 1). Within Hubei, the results generally agreed with the national pattern mentioned above. Daily-additional sick posts predicted daily-additional new cases in Hubei up to 19 days in advance, as illustrated in Figure 4B. In contrast, other COVID-19 posts had fewer lag terms that could forecast new cases. Outside Hubei, the predictive pattern of sick posts was similar to the national pattern despite a limited time range: sick posts could forecast new cases 2 to 8 days in advance (see Figure 4C).

Figure 4. Standardized estimates of Granger causality for time-lagged, daily-additional Weibo posts (sick posts and other COVID-19 posts) predicting daily-additional cases.



Discussion

Principal Findings

The novel coronavirus causing COVID-19 is a new pathogen in the human reservoir. It poses an extraordinary challenge for public health systems worldwide because screening and diagnostic tests must be developed from scratch. Even when such tests eventually become available, testing capacity is often severely limited; this can fuel the outbreak, as many patients can unknowingly infect other people. Based on approximately 15 million COVID-19-related Weibo posts between November 1, 2019 and March 31, 2020, we developed a supervised machine learning classifier to identify "sick posts," in which a user reports their own or other people's symptoms and diagnosis of COVID-19. Using the officially reported daily case counts as the outcome, our work shows that sick posts significantly predict daily cases up to 14 days ahead of official statistics. This finding confirms prior research that social media data can be usefully applied to nowcasting and forecasting emerging infectious diseases such as COVID-19 [22,34].

One of the greatest challenges of digital disease surveillance is identifying true disease signals, especially when facing the deluge of social media activity that resulted from COVID-19 mitigation measures [12,34-36]. Our finding that sick posts have greater predictive power than other COVID-19 posts shows that not all social media data are equally informative. Specifically, COVID-19 has dramatically disrupted everyday life; due to the pandemic, people are sheltering in place and increasingly communicating with others via social media. As shown in prior work [18] as well as in our data set, the majority of COVID-19–related chatter on Weibo reflected public awareness of COVID-19 rather than actual symptom reports. Most previous studies took rather coarse-grained approaches, relying primarily on either aggregated search query data or social media data retrieved from limited keyword searches [19,22]. In our work, we gathered the largest, most comprehensive, and most granular collection of social media data related to COVID-19 in the Chinese language. More importantly, we demonstrate a viable method to separate valid signals from noise using reports of symptoms and diagnosis, which makes a significant contribution to the literature on digital surveillance.

Another important finding is that while the predictive power of sick posts on daily case counts holds true for both Hubei and non-Hubei regions, the effect sizes vary. Being the epicenter of the outbreak, Hubei province experienced extreme testing shortages during the early stage of the study period. As a result, many Hubei residents turned to social media sites such as Weibo to seek help for testing and medical care. In contrast, social media help-seeking activities were uncommon in other parts of China, where testing and health care resources were much more adequate. Taking these regional variations into account, we still observed predictive signals of sick posts on case counts, suggesting that the predictive power of sick posts was robust against testing delays. Further, the variations in the effect estimates show that the predictive power of social media data may vary across different geographic areas, with different levels of preparedness, and at different stages of the outbreak. Future studies based on longer periods of data monitoring could explore the temporal and spatial variations of COVID-19 social media surveillance efficacy in more depth.

Our work has broad public health implications. The high speed and low cost of social media surveillance can be especially useful in the early stages of the COVID-19 outbreak to inform containment and mitigation efforts when they are most cost-effective. For countries and regions where public health infrastructures do not allow for widespread screening and diagnostic tests, social media disease surveillance provides much-needed information for public health agencies to model the trajectories of the outbreak and to make swift decisions about allocation of resources such as hospital beds, ventilators, and personal protective equipment.

Another advantage of social media surveillance is that it can be performed from a distance. As COVID-19 continues to spread worldwide, countries lacking testing and screening infrastructures will become "dark spots," endangering their own citizens as well as the entire world. It is imperative that international organizations such as the World Health Organization integrate such data into their outbreak forecasting management practices to mobilize and coordinate relief efforts to help combat COVID-19.

Limitations

This study has several limitations. First, Weibo posts were retrieved retrospectively rather than in real time; therefore, deleted or censored posts were absent from our data set. However, we have no reason to believe that deletion or censorship favored "sick posts" in measurable ways. In fact, a recent study on Weibo censorship from December 2019-February 2020 shows that only 1.7/1000 Weibo posts were censored; also, these censored posts generally pertained to the missteps in the government's COVID-19 response, not individual reports of symptoms and diagnoses [37]. Therefore, our results should not be affected by censorship. Second, as some studies suggest [38-40], confirmed COVID-19 case counts published by China CDC may underestimate the actual counts, due in part to limits in testing capacity and the existence of asymptomatic carriers. Still, the data here represent the best-known data of confirmed case counts, and our models rely on trends and changes in these case counts rather than the actual numbers. Third, it is important to acknowledge that sick posts as disease signals are not without noise because Weibo users who reported COVID-19 symptoms were not necessarily clinically diagnosed with COVID-19; Weibo users may not speak the truth; and Weibo users may "overreport" (posting about their symptoms or diagnoses multiple times) or "underreport" (not posting despite their symptoms or diagnoses) for a variety of reasons. Such inaccuracies are inherent in user-generated social media data and widely exist in all infoveillance studies. However, it should be noted that the goal of infoveillance has never been to achieve one-for-one matching between social media posts and clinical cases. Rather, infoveillance approaches strive to mine useful early signals from social media and internet data as a supplement to conventional surveillance efforts. Despite this noise, we still found that sick posts predicted COVID-19 case counts,

indicating the validity of this signal in reflecting disease spread in the population.

Conclusions

The threats of COVID-19 and other infectious diseases are likely to recur in the future. Reports of symptoms and diagnoses on social media during emerging disease outbreaks send invaluable warning signals to the public. Researchers and disease control agencies should pay close attention to the social media infosphere. In addition to monitoring overall search and posting activities, it is crucial to sift through the contents and efficiently separate true signals from noise. Our main findings highlight the importance of using rigorous procedures and understanding information sharing behaviors to obtain quality disease signals. Future studies based on longer periods of data monitoring could explore the time and spatial diffusions of COVID-19 in more depth. A more detailed examination of post contents reporting restraints in information or medical resources will be helpful in developing local outbreak responses.

Acknowledgments

We thank Jingyang Xu, Minwei Ren, Rixia Tang, Zichao Wang, Yongyan Xu, Na Yang, Yalan Jin, Xiuchan Xu, Xinyu Wang, Ruizhi Sun, Wenhui Zhu, Yiwei Li, and Tianyu Zhao for their help with data annotation.

Authors' Contributions

CS, WL, JZ, and BF contributed to the study design. AC collected the Weibo data. WL, CL and AC contributed to the data analysis. WL, CS, CL, and AC contributed to the design and drawing of the figures. All authors contributed to the writing of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables.
[DOCX File , 182 KB - jmir_v22i5e19421_app1.docx]

References

1. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb;395(10223):497-506. [doi: 10.1016/s0140-6736(20)30183-5]
2. Wu F, Zhao S, Yu B, Chen Y, Wang W, Song Z, et al. A new coronavirus associated with human respiratory disease in China. Nature 2020 Feb 3;579(7798):265-269. [doi: 10.1038/s41586-020-2008-3]
3. World Health Organization. 2020 May 16. Coronavirus disease 2019 (COVID-19) Situation Report 117 URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200516-covid-19-sitrep-117.pdf?sfvrsn=8f562cc_2 [accessed 2020-05-26]
4. Zhang J, Centola D. Social Networks and Health: New Developments in Diffusion, Online and Offline. Annu Rev Sociol 2019 Jul 30;45(1):91-109. [doi: 10.1146/annurev-soc-073117-041421]
5. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
6. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
7. Aiello AE, Renson A, Zivich PN. Social Media- and Internet-Based Disease Surveillance for Public Health. Annu Rev Public Health 2020 Apr 02;41:101-118. [doi: 10.1146/annurev-publhealth-040119-094402] [Medline: 31905322]
8. Barros JM, Duggan J, Rebholz-Schuhmann D. The Application of Internet-Based Sources for Public Health Surveillance (Infoveillance): Systematic Review. J Med Internet Res 2020 Mar 13;22(3):e13680 [FREE Full text] [doi: 10.2196/13680] [Medline: 32167477]
9. Charles-Smith LE, Reynolds TL, Cameron MA, Conway M, Lau EHY, Olsen JM, et al. Using Social Media for Actionable Disease Surveillance and Outbreak Management: A Systematic Literature Review. PLoS One 2015;10(10):e0139701 [FREE Full text] [doi: 10.1371/journal.pone.0139701] [Medline: 26437454]
10. Cui X, Yang N, Wang Z, Hu C, Zhu W, Li H, et al. Chinese social media analysis for disease surveillance. Pers Ubiquit Comput 2015 Sep 11;19(7):1125-1132. [doi: 10.1007/s00779-015-0877-5]
11. Fung IC, Fu K, Ying Y, Schaible B, Hao Y, Chan C, et al. Chinese social media reaction to the MERS-CoV and avian influenza A(H7N9) outbreaks. Infect Dis Poverty 2013 Dec 20;2(1):31 [FREE Full text] [doi: 10.1186/2049-9957-2-31] [Medline: 24359669]

12. Broniatowski DA, Paul MJ, Dredze M. National and local influenza surveillance through Twitter: an analysis of the 2012-2013 influenza epidemic. PLoS One 2013;8(12):e83672 [FREE Full text] [doi: 10.1371/journal.pone.0083672] [Medline: 24349542]
13. Klembczyk JJ, Jalalpour M, Levin S, Washington RE, Pines JM, Rothman RE, et al. Google Flu Trends Spatial Variability Validated Against Emergency Department Influenza-Related Visits. J Med Internet Res 2016 Jun 28;18(6):e175 [FREE Full text] [doi: 10.2196/jmir.5585] [Medline: 27354313]
14. Dugas AF, Hsieh Y, Levin SR, Pines JM, Mareiniss DP, Mohareb A, et al. Google Flu Trends: correlation with emergency department influenza rates and crowding metrics. Clin Infect Dis 2012 Feb 15;54(4):463-469 [FREE Full text] [doi: 10.1093/cid/cir883] [Medline: 22230244]
15. McGough SF, Brownstein JS, Hawkins JB, Santillana M. Forecasting Zika Incidence in the 2016 Latin America Outbreak Combining Traditional Disease Surveillance with Search, Social Media, and News Report Data. PLoS Negl Trop Dis 2017 Jan;11(1):e0005295 [FREE Full text] [doi: 10.1371/journal.pntd.0005295] [Medline: 28085877]
16. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. Science 2014 Mar 14;343(6176):1203-1205. [doi: 10.1126/science.1248506] [Medline: 24626916]
17. Li L, Zhang Q, Wang X, Zhang J, Wang T, Gao T, et al. Characterizing the Propagation of Situational Information in Social Media During COVID-19 Epidemic: A Case Study on Weibo. IEEE Trans Comput Soc Syst 2020 Apr;7(2):556-562. [doi: 10.1109/tcss.2020.2980007]
18. Zhu Y, Fu K, Grépin KA, Liang H, Fung IC. Limited Early Warnings and Public Attention to Coronavirus Disease 2019 in China, January-February, 2020: A Longitudinal Cohort of Randomly Sampled Weibo Users. Disaster Med Public Health Prep 2020 Apr 03:1-4 [FREE Full text] [doi: 10.1017/dmp.2020.68] [Medline: 32241328]
19. Li J, Xu Q, Cuomo R, Purushothaman V, Mackey T. Data Mining and Content Analysis of the Chinese Social Media Platform Weibo During the Early COVID-19 Outbreak: Retrospective Observational Infoveillance Study. JMIR Public Health Surveill 2020 Apr 21;6(2):e18700 [FREE Full text] [doi: 10.2196/18700] [Medline: 32293582]
20. Zhao Y, Cheng S, Yu X, Xu H. Chinese Public's Attention to the COVID-19 Epidemic on Social Media: Observational Descriptive Study. J Med Internet Res 2020 May 04;22(5):e18825 [FREE Full text] [doi: 10.2196/18825] [Medline: 32314976]
21. Huang C, Xu X, Cai Y, Ge Q, Zeng G, Li X, et al. Mining the Characteristics of COVID-19 Patients in China: Analysis of Social Media Posts. J Med Internet Res 2020 May 17;22(5):e19087 [FREE Full text] [doi: 10.2196/19087] [Medline: 32401210]
22. Li C, Chen LJ, Chen X, Zhang M, Pang CP, Chen H. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. Euro Surveill 2020 Mar;25(10):1 [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.10.2000199] [Medline: 32183935]
23. weibo.com. 2019 Annual Sina Weibo User Report URL: http://ir.weibo.com/node/7726/html
24. Chen Z, Su CC, Chen A. Top-down or Bottom-up? A Network Agenda-setting Study of Chinese Nationalism on Social Media. J Broadcasting Electron Media 2019 Sep 20;63(3):512-533. [doi: 10.1080/08838151.2019.1653104]
25. Li Y, Luo C, Chen A. The evolution of online discussions about GMOs in China over the past decade: Changes, causes and characteristics. Cultures of Science 2020 Jan 20;2(4):311-325. [doi: 10.1177/209660831900200406]
26. weibo.com. 2019. 2018 Annual Sina Weibo User Report. Webpage in Chinese URL: https://data.weibo.com/report/reportDetail?id=433
27. Lacy S, Watson BR, Riffe D, Lovejoy J. Issues and Best Practices in Content Analysis. Journal Mass Commun Q 2015 Sep 28;92(4):791-811. [doi: 10.1177/1077699015607338]
28. Hu Y, Huang H, Chen A, Mao XL. arXiv. 2020 May 21. Weibo-COV: A Large-Scale COVID-19 Social Media Dataset from Weibo URL: https://arxiv.org/abs/2005.09174 [accessed 2020-05-26]
29. Sun K, Chen J, Viboud C. Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study. Lancet Digit Health 2020 Apr;2(4):e201-e208 [FREE Full text] [doi: 10.1016/S2589-7500(20)30026-1] [Medline: 32309796]
30. China Center for Disease Control and Prevention. 2020 Feb 12. COVID-19 Situation Report on Feburary 12, 2020. Webpage in Chinese URL: http://www.chinacdc.cn/jkzt/crb/zl/szkb_11803/jszl_11809/202002/t20200213_212624.html [accessed 2020-05-26]
31. Granger CWJ. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. Econometrica 1969 Aug;37(3):424. [doi: 10.2307/1912791]
32. Box GEP, Tiao GC. Intervention Analysis with Applications to Economic and Environmental Problems. J Am Stat Assoc 1975 Mar;70(349):70-79. [doi: 10.1080/01621459.1975.10480264]
33. Box-Steffensmeier J, Freeman J, Hitt M, Pevehouse J. Time Series Analysis for the Social Sciences. Cambridge, UK: Cambridge University Press; 2014.
34. Buckee C. Improving epidemic surveillance and response: big data is dead, long live big data. Lancet Digit Health 2020 May;2(5):e218-e220 [FREE Full text] [doi: 10.1016/s2589-7500(20)30059-5]
35. Hua J, Shaw R. Corona Virus (COVID-19) "Infodemic" and Emerging Issues through a Data Lens: The Case of China. Int J Environ Res Public Health 2020 Mar 30;17(7):2309 [FREE Full text] [doi: 10.3390/ijerph17072309] [Medline: 32235433]

36. Leung GM, Leung K. Crowdsourcing data to mitigate epidemics. Lancet Digit Health 2020 Apr;2(4):e156-e157. [doi: 10.1016/s2589-7500(20)30055-8]
37. Fu K, Zhu Y. Did the world overlook the media's early warning of COVID-19? J Risk Res 2020 Apr 24:1-5. [doi: 10.1080/13669877.2020.1756380]
38. Kucharski A, Russell T, Diamond C, Liu Y, Edmunds J, Funk S, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. Lancet Infect Dis 2020 May;20(5):553-558 [FREE Full text] [doi: 10.1016/S1473-3099(20)30144-4]
39. Imai N, Dorigatti I, Cori A, Donnelly C, Riley S, Ferguson N. Imperial College. 2020 Jan 22. Report 2: Estimating the potential total number of novel Coronavirus (2019-nCoV) cases in Wuhan City, China URL: https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-update-epidemic-size-22-01-2020.pdf [accessed 2020-05-26]
40. Wu J, Leung K, Leung G. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet 2020 Feb;395(10225):689-697. [doi: 10.1016/S0140-6736(20)30260-9]

Abbreviations

API: application programming interface
China CDC: Chinese Center for Disease Control and Prevention
COVID-19: coronavirus disease

Edited by G Eysenbach; submitted 16.04.20; peer-reviewed by H Liang, KW Fu, E Lau, C Basch; comments to author 08.05.20; revised version received 18.05.20; accepted 25.05.20; published 28.05.20.

Please cite as:
Shen C, Chen A, Luo C, Zhang J, Feng B, Liao W
Using Reports of Symptoms and Diagnoses on Social Media to Predict COVID-19 Case Counts in Mainland China: Observational Infoveillance Study
J Med Internet Res 2020;22(5):e19421
URL: http://www.jmir.org/2020/5/e19421/
doi:10.2196/19421
PMID:32452804

Original Paper

Early Detection of Mild Cognitive Impairment With In-Home Sensors to Monitor Behavior Patterns in Community-Dwelling Senior Citizens in Singapore: Cross-Sectional Feasibility Study

Iris Rawtaer[1], MBBS, MMed, MCI; Rathi Mahendran[2], MBBS, MMed, FAMS; Ee Heok Kua[2], MBBS, MD, FRCPsych; Hwee Pink Tan[3], PhD; Hwee Xian Tan[3], PhD; Tih-Shih Lee[4], MD, PhD; Tze Pin Ng[2], MBBS, PhD

[1]Department of Psychiatry, Sengkang General Hospital, Singhealth Duke NUS Academic Medical Centre, Singapore, Singapore

[2]Department of Psychological Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[3]School of Information Systems, Singapore Management University, Singapore, Singapore

[4]Duke-NUS Graduate Medical School, Singapore, Singapore

Corresponding Author:
Iris Rawtaer, MBBS, MMed, MCI
Department of Psychiatry
Sengkang General Hospital
Singhealth Duke NUS Academic Medical Centre
110 Sengkang East Way
Singapore, 544886
Singapore
Phone: 65 69302288
Email: iris.rawtaer@singhealth.com.sg

Abstract

Background: Dementia is a global epidemic and incurs substantial burden on the affected families and the health care system. A window of opportunity for intervention is the predementia stage known as mild cognitive impairment (MCI). Individuals often present to services late in the course of their disease and more needs to be done for early detection; sensor technology is a potential method for detection.

Objective: The aim of this cross-sectional study was to establish the feasibility and acceptability of utilizing sensors in the homes of senior citizens to detect changes in behaviors unobtrusively.

Methods: We recruited 59 community-dwelling seniors (aged >65 years who live alone) with and without MCI and observed them over the course of 2 months. The frequency of forgetfulness was monitored by tagging personal items and tracking missed doses of medication. Activities such as step count, time spent away from home, television use, sleep duration, and quality were tracked with passive infrared motion sensors, smart plugs, bed sensors, and a wearable activity band. Measures of cognition, depression, sleep, and social connectedness were also administered.

Results: Of the 49 participants who completed the study, 28 had MCI and 21 had healthy cognition (HC). Frequencies of various sensor-derived behavior metrics were computed and compared between MCI and HC groups. MCI participants were less active than their HC counterparts and had more sleep interruptions per night. MCI participants had forgotten their medications more times per month compared with HC participants. The sensor system was acceptable to over 80% (40/49) of study participants, with many requesting for permanent installation of the system.

Conclusions: We demonstrated that it was both feasible and acceptable to set up these sensors in the community and unobtrusively collect data. Further studies evaluating such digital biomarkers in the homes in the community are needed to improve the ecological validity of sensor technology. We need to refine the system to yield more clinically impactful information.

(J Med Internet Res 2020;22(5):e16854) doi:10.2196/16854

KEYWORDS

dementia; neurocognitive disorder; pattern recognition, automated/methods; internet of things; early diagnosis

Introduction

Background

Dementia is a neurodegenerative disease of epidemic proportions and incurs substantial burden on the affected families and the health care system. Globally, about 47 million people were living with dementia in 2015, and this number is projected to triple in 2050. The global costs of dementia in 2015 were estimated at US $818 billion, a 35.4% increase compared with 2010. Up to 1 in 5 of the community-dwelling older adults aged 65 years and above suffer from mild cognitive impairment (MCI) [1]. MCI is considered as an at-risk state for dementia, a stage when the affected person is likely responsive to appropriate interventions. Existing evidence suggests that multidomain, multicomponent interventions can improve or maintain cognitive function of persons with MCI, delaying further cognitive decline [2]. If dementia onset is delayed by 5 years, then it would halve the global dementia prevalence and would result in substantial reduction in medical, family, and social care burden for dementia. There is an urgent need for the early detection of MCI to facilitate monitoring and intervention and to allow individuals and their families to plan ahead. According to the practice guidelines for MCI by the American Academy of Neurology [3], detecting MCI early, performing serial assessments and implementing interventions, and allowing individuals to plan ahead are essential.

Delayed recognition of MCI in the community is a missed opportunity for early intervention. One challenge that clinicians face is the reliance on patients or their relatives for reporting subtle changes or decline, and these often occur when the cognitive decline is relatively advanced. As a result, only a fraction of individuals with MCI seek early medical attention. There are many undiagnosed MCI cases in the community who would only seek help when they have more fulminant symptoms of dementia, thus missing the opportunity of having early interventions to delay the progression to dementia. In rapidly ageing societies, working adult children cannot be relied upon to detect these subtle changes; hence, innovative methods have to be employed to be our *eyes and ears* in the community.

Technology can be these *eyes and ears*. Indeed, sensor technology is quickly gaining popularity in the medical community for its utility in continuous health monitoring [4]. In contrast to the extensive literature on the usage of sensors to detect falls, assess gait, and remotely monitor physical health [5-7], the use of sensors to monitor cognition and mental well-being is comparatively less well studied. Extant studies are often laboratory based and conducted in test beds or facilities [4,8,9]. A systematic review of home-based monitoring of cognitive function published this year indicated that few studies have done real-life evaluations in uncontrolled conditions [10]. Some earlier studies looked at continuous recording of daily audio patterns and linked it to social and mental well-being [11], whereas others looked at utilizing mobile phone sensors to evaluate depressive symptom severity and physiological signals associated with mental, emotional, and physical stresses [12,13]. Specific to cognition, one group approached early detection of dementia by using infrared sensors to monitor in-house activities; they found that subjects with impaired cognition had lesser number of outings and a shorter sleep time compared with controls [14]. This same group replicated their findings in a bigger sample over the course of 1 year and found that senior citizens who had shown cognitive decline had lesser outings [15]. Similar passive in-home sensor setups were reported in other two studies evaluating cognitive states [16,17]. The coefficient of variation of median walking speeds in the MCI group was twice that of controls [16], and there were distinct trajectories of walking speeds between different cognitive states over 2.6 (SD 1.0) years of follow-up [17].

To establish feasibility and acceptability of remote monitoring of senior citizens' behavior patterns in our community, we conducted a pilot study utilizing multiple sensors installed at home to capture certain behaviors. These behaviors would typically be assessed in a clinical evaluation, such as forgetfulness, sleep, and activity levels. We hypothesized that (1) sensor-derived data of specific activity patterns between 2 groups of community-dwelling seniors, those with MCI and those who are cognitively healthy (healthy cognition; HC), will differ and (2) sensors for remote monitoring in the homes of senior citizens would be acceptable.

Methods

Study Design and Participant Recruitment

This was a cross-sectional study conducted over a period of 2 months. The study commenced in March 2016 and was completed in August 2018. Participants were recruited from existing community studies such as the Singapore Longitudinal Ageing Study and Jurong Ageing Study (JAS). These participants had previously consented to be recontacted for related studies. Participants were also recruited from the community through senior citizen activity centers such as Presbyterian Community Services. Institutional ethics review board approvals were obtained (reference number: 2015/01076).

Informed consent was obtained before participants were screened for eligibility. Participants were included if they were (1) aged between 65 and 85 years, (2) living alone, (3) able to provide written informed consent in English/Mandarin, and (4) available for the entire duration of the study. Participants were excluded if they (1) had a previous diagnosis of dementia/any neurodegenerative condition, (2) had a diagnosis of any psychiatric disorder, (3) had limitations of physical mobility or required assistance with their activities of daily living, or (4) were not willing to have sensors deployed in stipulated areas of the home.

Data Collection

At baseline, basic sociodemographic data were collected. Depressive symptoms were elicited using the Zung Self-Rating Depression Scale (SDS) and Geriatric Depression Scale (GDS). Subjective sleep quality was captured using the Pittsburgh Sleep Quality Index (PSQI). The Friendship Scale (FS) was used to capture social connectedness. Measures of global cognition including the Montreal Cognitive Assessment (MoCA) and modified Mini-Mental State Examination (MMSE) were administered. Participants who had not been administered the

Clinical Dementia Rating (CDR) or neuropsychological test batteries within the past 6 months (as part of other studies) performed these tests. Structured Clinical Interview for DSM Disorders was administered to all participants. Participants who screened positive for any DSM disorder were excluded. SDS, GDS, PSQI, FS, MMSE, and MoCA were repeated at 2 months. Participants were asked for feedback at the end of the 2 months of participation. This was an unstructured qualitative written feedback.

Participants' cognitive status of HC vs MCI was established at baseline with the MMSE, MoCA, CDR, and neuropsychological test performance and via a consensus panel.

The *diagnosis of MCI* was defined according to the following published criteria: (1) subjective memory and cognitive difficulties or informant/clinician-observed cognitive difficulties; (2) objective cognitive impairment in one or more domains: MMSE global score ranging from 24 to 27 and at least one neurocognitive domain (attention, memory, executive function, language, or visuospatial abilities) with a score of 1 to 2 standard deviations less than the age- and education-adjusted mean values; (3) CDR scale global score >0.5; (4) essentially independent in performing basic activities of daily living; and (5) not demented. Subtyping into amnestic and nonamnestic MCI based on the presence or absence of memory impairment was done according to established criteria [18].

Sensor Setup and Behaviors Captured

Upon completion of baseline assessments, participants had their homes instrumented with a network of sensors (Figure 1) for a duration of 2 months. The multimodal sensor system comprised passive infrared (PIR) motion sensors, proximity beacon tags, a sensor-equipped medication box (forgetfulness; Figure 2), a bed sensor (sleep), and a wearable (pedometer and heart rate). Each sensor periodically sensed the physical environment and then wirelessly transmitted the sensed data to the gateway. The gateway transmitted the aggregated data to the backend server via secure cellular communications (eg, 3G) for monitoring and processing. Each data point was identifiable only via the sensor node identifier; the mapping between the sensor node identifier and the home was securely stored and accessible only by the study investigators.

The sensor network was used to capture several behaviors of interest. The main feature/outcome of interest was *forgetfulness*. The other features were in-home activity levels, sleep quality, and physical activity; other changes are seen in cognitive decline but are often overlooked.

A combination of sensor data was used to measure forgetfulness. Participants were provided with a sensor-equipped medication box to store all their prescription medication; data were generated whenever the box was opened [19]. These data, taken together with the expected medication frequency information obtained at baseline, allowed us to determine the number of times a participant forgot to take their medication at the prescribed time. Proximity beacon tags were attached to participants' personal effects such as keychains and wallet, allowing us to estimate the distance between the item and the home gateway. Efforts were made at baseline to ensure that these personal effects were items that were routinely brought out. Coupled with the wearable and PIR motion sensors, we were able to determine if the participant had forgotten to bring these items with them when leaving home. The faucet usage sensor was used to determine if the participant had forgotten to switch off the faucet after moving away from the designated area (detected by motion sensors). The in-home activity levels and number of outings were inferred from the PIR motion sensors and door contact sensor, which detects opening and closing of the main door of the participant's residence. The bed sensor (based on fiberoptic technology) placed under the participant's mattress provided data on sleep duration and quality. The wearable activity band (Microsoft band) measured heart rate and daily steps. Participants were expected to wear the band at all times, with the exception of shower time. The smart plug was used to detect if specific appliances in the home were used, most commonly the television [20].

Figure 1. Multimodal sensor set up in homes.

Figure 2. Proximity beacon tags to wallet and keys and sensor-equipped medication box.

Safety

Although detection of no movement was not an aim of this study, given that it is a capability of the system, we provided this feature for safety monitoring as all the senior citizens enrolled were living alone. As such, if the system detected no movement for 8 hours, an alert was sent to the caregiver, care provider (social service agency), or research team.

Sample Size Justification

An earlier study with a sample size of 14 comparing walking speeds and variance in MCI and healthy controls using continuously obtained data from sensors revealed a significant difference [16]. Given the feasibility and resource constraints of the study, we planned a priori to enroll 36 participants at a minimum.

Data Analysis

Raw sensor data readings were converted into a common format and aggregated into a database. Purging of the data was performed to remove erroneous data and periods where the system was down/partially down. It was necessary to remove days where the system was down, as it would affect the *forgetfulness* metric. After data purging, sensor-specific data cleaning or validation was performed to ensure that only valid sensor data are processed. Frequency of each metric was computed, for example, frequency of forgetting medication per month, frequency of outings per day, and frequency of sleep interruptions per night. Investigators involved in obtaining sensor data and analysis were blinded to participants' cognitive status.

Descriptive statistics for demographic information and psychometric scores were computed for all participants at

baseline. As psychometric measures were administered at baseline and at 2 months, these scores were averaged to provide a precise cross-sectional estimate. Demographic, psychometric characteristics, and behavior metrics (computed from sensor data) were compared between the HC participants and MCI participants using the Student *t* test or the Wilcoxon rank sum test for continuous variables and the Pearson chi-square test and the Fisher exact test for categorical variables. Depending on the distribution of data, Pearson or Spearman tests were applied to look at a correlation between sensor-derived data and psychometric test measures.

Results

Participant Characteristics

A total of 59 participants were screened. One participant was ineligible because of an existing mental health condition. Another 8 participants were not enrolled because of reasons of hospitalization, family objections, or overseas travel during the study period. One participant was enrolled but dropped out shortly after the sensors were deployed because of discomfort with the bed sensor. In total, 49 participants completed the study. Of the 49 participants, 28 were diagnosed with MCI and 21 with HC. Half of the MCI participants were of the amnestic subtype and half were of the nonamnestic subtype. Participant demographics and psychometric measures are shown by group in Table 1. There were no statistically significant differences in demographics between the completers (n=49) and the noncompleters (n=10).

Table 1. Participant characteristics by group—normal cognition vs mild cognitive impairment.

Demographics	Cognitively healthy (n=21)	Mild cognitive impairment (n=28)
Age (years), mean (SD)	73.0 (5.3)	75.1 (6.3)
Gender, n (%)		
Male	7 (33)	9 (32)
Female	14 (67)	19 (68)
Years of education, mean (SD)	7.0 (4.0)	4.5 (3.9)
Employment (currently employed part time), n (%)	8 (38)	2 (7)
Housing type—Housing Development Board 1-2 room flat, n (%)	14 (67)	20 (71)
Marital status, n (%)		
Never married	8 (38)	10 (36)
Separated/divorced	7 (33)	9 (32)
Widowed	5 (24)	9 (32)
Medical conditions, n (%)		
Hypertension	11 (52)	17 (60)
Hyperlipidemia	12 (57)	19 (68)
Diabetes mellitus	4 (19)	6 (22)
Stroke	0 (0)	3 (11)
Ischemic heart disease	2 (10)	4 (14)
Psychometric measures, mean (SD)		
Mini-Mental State Examination	28.1 (3.2)	26.3 (2.2)
Montreal Cognitive Assessment	27.5 (1.6)	24.0 (3.1)
Geriatric Depression Scale	0.6 (0.7)	1.4 (1.0)
Zung Depression Scale	44.5 (2.1)	42.7 (3.4)
Pittsburg Sleep Quality Index	3.8 (3.2)	5.0 (2.2)
Friendship Scale	18.5 (1.7)	19.1 (1.7)

Behaviors of Interest and Psychometric Measures

Frequencies of incidents of forgetfulness and various behaviors of interest (as described above) were computed and compared between the MCI and HC groups (Table 2).

As expected, the MCI group had lower MMSE scores than the HC group. They also tended to have poorer sleep quality, with higher scores on the PSQI. With the sensor-derived data, we found that MCI participants were less active than their HC counterparts; MCI participants had an average of 3407 steps a

day compared with 4033 steps in the HC group. MCI participants spent less time away from home daily. They had more sleep interruptions per night (2 per night) compared with the HC group (1 per night). The MCI group had forgotten their medications an average of two times more per month compared with the HC group (30 vs 28). The MCI group forgot their wallet when leaving the home at a similar frequency to the HC group. Unexpectedly, the HC group had a higher frequency of forgetting their keys per month as compared with the MCI group. It is important to note that none of these differences achieved statistical significance. Faucet use data were not analyzed as the sample size of usable data was too small; there were many implementation issues with the sensor. Correlation analysis of sensor-derived behavior metrics with psychometric measures did not yield any significant results.

Table 2. Comparisons of activities between the cognitively healthy group and mild cognitive impairment group.

Behaviors of interest, mean (SD)	Cognitively healthy (n=21)	Mild cognitive impairment (n=28)	*P* value
Steps (daily)	4033 (2148)	3407 (2688)	.40
Heart rate (bpm)	72 (4)	75 (7)	.22
Sleep duration daily (min)	440 (155)	427 (246)	.84
Number of sleep interruptions	1 (1)	2 (2)	.27
Number of outings daily	1 (1)	1 (1)	.93
Time away from home daily (min)	300 (153)	267 (132)	.44
Frequency of forgetting medication/month	28 (13)	30 (28)	.85
Frequency of forgetting keys per month	21 (16)	17 (13)	.40
Frequency of forgetting wallet per month	24 (17)	24 (22)	.94
Television use daily (min)	174 (176)	219 (220)	.52

Acceptability

A total of 83% (41/49) of the participants gave positive feedback at the study conclusion. Many participants found it reassuring that the system was able to detect deviations in their daily activity patterns and liked it that someone was aware if they declined physically or psychologically. Many asked if there was an option for the system to be permanently installed. Some negative feedback included needing to charge the wearable, worrying about electricity costs, and finding the water usage sensor inconvenient. Selected positive and negative feedback obtained from the participants at study conclusion are shown in Textbox 1.

Textbox 1. Selected positive and negative written feedback from the participants about the system.

- "The sensor system is very good for me who is weak and old and living alone. In case anything happens to me, I know someone will help me, feel very secure and safe. The staff who came to install the sensor are very friendly, kind and helpful. The watch does not cause any inconvenience to me all. Instead, I feel very safe to go out alone knowing that my movements are being monitored. It will be better if I have the sensors installed at my place permanently."
- "(1) With the sensors, feel secure - can sleep better. (2) Not used to the blinking light. (3) Found trouble in wearing the sensor watch. (4) Feel uncomfortable with the gadget near bathroom."
- "The sensor system is very good for senior citizens living alone, feel very secure knowing my movements are being monitored all the time. The watch is very good. I can go out anytime. It will record my movements. One of my old neighbours was staying upstairs alone. The lady died in her toilet but her body was discovered many days later. If only she had the sensor system installed then she will be able to get help earlier."
- "System is unobtrusive. Sensors are small and do not inconvenience me in any way. Enjoy the use of the Microsoft band. Does not alter her activities because of sensors. Bed sensor does not disturb her and she thinks she sleeps as per her sleep patterns. Thinks it is a useful system for elderly who live alone."
- "In home sensors is very good for senior citizens living alone. In case anything happens to me, I am able to get help. Feel secure with sensors installed. The watch is good, can monitor my steps. Feel safe to go out alone. Disadvantage: - Electricity bill increased a little. Too many wires that take up space."
- "In-home sensor is very helpful for senior citizens living alone. Multiple sensors installed taking up some of the space in the house. The one near the kitchen sink interfere with my daily washing, unable to clean properly, otherwise no other problem."
- "I feel very safe and relax with sensors on. If possible, I would like to have the sensors installed in my house permanently. The whole system does not cause any problems to me at all."
- "On a whole, quite satisfied with the sensors installed except I don't like one of the sensors installed near my kitchen sink, one of the tubing extending halfway into the washing bowl (sink) causing inconvenience when washing my kitchen utensils. Hope this project able to help those elderly living alone in future."
- "(1) Very satisfied with the sensors installed. (2) Like the watch as it can monitor my steps. (3) Does not cause any inconvenience. (4) Feel more secure and comfortable. (5) Overall, I feel good and happy and even request for extension. (6) Hope to have sensors installed permanently."
- "I found sensors are very good. Every day when I came back, sensor lights will flicker, and I will feel very safe."
- "Basically, it's good to be monitored for safety reasons. I don't find it a nuisance, but electricity is left on for 2 months."
- "(1) Sensor is OK. (2) A bit of inconvenience with the sensor at the sink. (3) I seldom wear the watch when I go out."
- "Good, not interfering with daily living but use more electricity."
- "not suitable for me; suitable for blur and not very smart people; troublesome to wear the watch all the time; recommend for those very forgetful people"
- "Satisfied with the gadgets. Helpful for monitoring."
- "Scared of the gadget with all the plugs and lighting."
- "I feel very safe and relax with the sensors installed in my home; I also feel safe to go out wearing the watch; the sensor gadget does not interfere with my daily activities; good for elderly living alone."
- "sensor system is good; monitor any movements at home to detect anything unusual; the watch is also good to monitor number of steps and movements outside my home and I can go out as usual with no restriction; no problems with all the sensor gadgets installed at home; highly recommended for elderly living alone."

Discussion

Principal Findings

The results of this pilot study indicate that it is feasible to set up sensor networks in the homes of community-dwelling seniors and unobtrusively collect potentially meaningful clinical data. Differences between HC and MCI in several behaviors of interest including daily activity (step count and time away from home), sleep (duration and interruptions), and forgetting medication suggest that cross-sectional remote observation of behaviors can yield discernible patterns, albeit not achieving statistical significance. Our observations on activity and sleep measures are consistent with earlier studies [14,15] and existing understanding of MCI [21,22]. Some collected sensor-derived data were counter-intuitive—minimal differences in frequency of forgetting wallet and higher frequency of forgetting keys in HC. A possible explanation for these observations may be that the study period of 2 months was short; hence, lower-frequency events are less likely to demonstrate identifiable patterns. Half of the MCI participants were of the nonamnestic subtype; hence, they may not have demonstrated differences in the forgetfulness behavior metric. With a longer observation period, behavioral markers that are better indicators of cognitive impairment can be determined.

Given our initial concerns that senior citizens would be wary of remote monitoring systems and past research on unobtrusive systems indicating privacy or security concerns [23], over 80% of the participants giving positive feedback and finding the system acceptable was an immensely encouraging result. Negative feedback was related to practical user issues with

specific devices such as having to charge the wearable, blinking lights of the motion sensor, and the need for multiple plugs. Despite the largely positive reception of the system from participants, we are careful not to extrapolate this acceptability to all senior citizens in general. The senior citizens in this study were all living alone and over two-thirds lived in 1-2 room public housing, a proxy of lower socioeconomic status. They are a group most at risk of undetected cognitive decline and most likely to benefit from an unobtrusive home monitoring system. Many existing sensor systems on trial by social care agencies for this group are specifically set up to detect falls or death, outcomes that sometimes go unnoticed in this group of vulnerable senior citizens.

Strengths and Limitations

Strengths of this pilot study include the use of an entirely unobtrusive system without the use of cameras, protecting the privacy of participants. Moreover, clinically useful behavior metrics such as forgetfulness, activity levels, and sleep were captured. These sensors were trialed in actual residential homes and not in laboratory settings or assisted living facilities; senior citizens were observed in their natural environments with no change to their lifestyle. Unobtrusive observation in the senior citizens' own homes highlights the potential translational value of having technology be the *eyes and ears* to monitor the senior citizens' health without taxing the working adult population.

This study was planned to evaluate feasibility and was exploratory in nature. It is limited by its small sample size and short duration of observation. There was some data loss in the initial phases of the study when the system was down. This was quickly addressed, and the system uptime was at 80% to 90% for much of the study duration. The accuracy of some of the behavior metrics that act as proxies for *forgetfulness* will need to be refined in future studies. Missed medication doses may indicate both the intentional nonadherence to prescribed doses of medication or genuine forgetfulness. In future studies, apart from capturing baseline medication intake frequency, there should be a measure of baseline adherence to various medications. Similarly, the behavior metric of forgetting personal effects will also require fine-tuning in subsequent studies. Forgetting to switch off the faucet is a commonly asked clinical question. Unfortunately, the technical trial of the water usage sensor was not successful in this study. Finally, participants were recruited from previous cohort studies as well as senior citizen activity centers, and this may have led to selection bias. Moreover, the inability of motion sensors to differentiate between 2 unique individuals led to an inherent selection bias, with only individuals living alone being recruited.

Conclusions

We found that it was both feasible and acceptable to use sensors to unobtrusively monitor behavior patterns in the homes of community-dwelling senior citizens. A larger study with a longer observation period of 2 to 3 years is underway. Negative feedback from the pilot has been addressed as far as practicable, including using fewer sensors and a different wearable. Analysis of trajectories and variability over time will yield more useful information, including behavior patterns that predict decline from MCI to dementia. Artificial intelligence methods including supervised learning models will be applied. Moving forward, we need to look at reducing the number of sensors to obtain more information. To improve generalizability to all senior citizens rather than only those who live alone, innovative solutions are needed to circumvent limitations of the motion sensor while still maintaining privacy. Identifying sensor-derived behavior metrics that do away with the motion sensor will allow the inclusion of senior citizens who live with others, improving the scalability of this solution. Although the system does not comprise video cameras, addressing privacy and security concerns is paramount when implementing and refining remote monitoring systems. Utilizing Internet of Things and artificial intelligence to monitor cognitive and physical well-being should be further developed to deliver value-added health care for senior citizens. Early detection of anomalies allows for self-management, timely interventions in the home and community, and facilitating remote capture of clinically meaningful data that can be utilized by health care professionals for diagnostic and prognostic purposes.

Acknowledgments

This study was funded by grants from the National University of Singapore's Clinician Scientist Unit Pitch for Funds Round (2016) and the National Medical Research Council's Healthcare Research Scholarship (2015). In-kind sponsorship of some devices was provided by Singapore Management University. The funders of the study had no role in the study design, data collection, data analysis, data interpretation, or writing of the manuscript. The corresponding author had full access to all the data and the final responsibility to submit for publication. The study team would like to acknowledge Presbyterian Community Services and Thye Hua Kwan Moral Charities for their kind support. The authors thank the Singapore Longitudinal Ageing Study and JAS teams.

Authors' Contributions

IR planned the study, conducted the literature search, collected data, performed data analysis, contributed to data interpretation, and drafted the manuscript. NT and RM contributed to study design, data collection, data analysis, data interpretation, and manuscript writing. TH and TH contributed to data collection (sensors), data analysis, and manuscript writing. KE and LT contributed to data analysis, data interpretation, and manuscript writing.

Conflicts of Interest

None declared.

References

1. Livingston G, Sommerlad A, Orgeta V, Costafreda SG, Huntley J, Ames D, et al. Dementia prevention, intervention, and care. Lancet 2017 Dec 16;390(10113):2673-2734. [doi: 10.1016/S0140-6736(17)31363-6] [Medline: 28735855]
2. Ngandu T, Lehtisalo J, Solomon A, Levälahti E, Ahtiluoto S, Antikainen R, et al. A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial. Lancet 2015 Jun 6;385(9984):2255-2263. [doi: 10.1016/S0140-6736(15)60461-5] [Medline: 25771249]
3. Petersen RC, Lopez O, Armstrong MJ, Getchius TS, Ganguli M, Gloss D, et al. Practice guideline update summary: Mild cognitive impairment: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology. Neurology 2018 Jan 16;90(3):126-135 [FREE Full text] [doi: 10.1212/WNL.0000000000004826] [Medline: 29282327]
4. Rantz MJ, Skubic M, Miller SJ, Galambos C, Alexander G, Keller J, et al. Sensor technology to support aging in place. J Am Med Dir Assoc 2013 Jun;14(6):386-391 [FREE Full text] [doi: 10.1016/j.jamda.2013.02.018] [Medline: 23562281]
5. Reeder B, Meyer E, Lazar A, Chaudhuri S, Thompson HJ, Demiris G. Framing the evidence for health smart homes and home-based consumer health technologies as a public health intervention for independent aging: a systematic review. Int J Med Inform 2013 Jul;82(7):565-579 [FREE Full text] [doi: 10.1016/j.ijmedinf.2013.03.007] [Medline: 23639263]
6. Patel S, Park H, Bonato P, Chan L, Rodgers M. A review of wearable sensors and systems with application in rehabilitation. J Neuroeng Rehabil 2012 Apr 20;9:21 [FREE Full text] [doi: 10.1186/1743-0003-9-21] [Medline: 22520559]
7. Taraldsen K, Chastin SF, Riphagen II, Vereijken B, Helbostad JL. Physical activity monitoring by use of accelerometer-based body-worn sensors in older adults: a systematic literature review of current knowledge and applications. Maturitas 2012 Jan;71(1):13-19. [doi: 10.1016/j.maturitas.2011.11.003] [Medline: 22134002]
8. Dawadi PN, Cook DJ, Schmitter-Edgecombe M, Parsey C. Automated assessment of cognitive health using smart home technologies. Technol Health Care 2013;21(4):323-343 [FREE Full text] [doi: 10.3233/THC-130734] [Medline: 23949177]
9. Glenn T, Monteith S. New measures of mental state and behavior based on data collected from sensors, smartphones, and the Internet. Curr Psychiatry Rep 2014 Dec;16(12):523. [doi: 10.1007/s11920-014-0523-3] [Medline: 25308392]
10. Piau A, Wild K, Mattek N, Kaye J. Current state of digital biomarker technologies for real-life, home-based monitoring of cognitive function for mild cognitive impairment to mild Alzheimer disease and implications for clinical care: systematic review. J Med Internet Res 2019 Aug 30;21(8):e12785 [FREE Full text] [doi: 10.2196/12785] [Medline: 31471958]
11. Rabbi M, Ali S, Choudhury T, Berke E. Passive and In-Situ Assessment of Mental and Physical Well-being Using Mobile Sensors. In: Proceedings of the 13th international conference on Ubiquitous computing. 2011 Presented at: UbiComp'11; September 17 - 21, 2011; Beijing, China p. 385-394 URL: http://europepmc.org/abstract/MED/25285324 [doi: 10.1145/2030112.2030164]
12. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. J Med Internet Res 2015 Jul 15;17(7):e175 [FREE Full text] [doi: 10.2196/jmir.4273] [Medline: 26180009]
13. Mohino-Herranz I, Gil-Pita R, Ferreira J, Rosa-Zurera M, Seoane F. Assessment of mental, emotional and physical stress through analysis of physiological signals using smartphones. Sensors (Basel) 2015 Oct 8;15(10):25607-25627 [FREE Full text] [doi: 10.3390/s151025607] [Medline: 26457710]
14. Suzuki T, Murase S, Tanaka T, Okazawa T. New approach for the early detection of dementia by recording in-house activities. Telemed J E Health 2007 Feb;13(1):41-44. [doi: 10.1089/tmj.2006.0033] [Medline: 17309353]
15. Suzuki T, Murase S. Influence of outdoor activity and indoor activity on cognition decline: use of an infrared sensor to measure activity. Telemed J E Health 2010;16(6):686-690. [doi: 10.1089/tmj.2009.0175] [Medline: 20575611]
16. Hayes TL, Abendroth F, Adami A, Pavel M, Zitzelberger TA, Kaye JA. Unobtrusive assessment of activity patterns associated with mild cognitive impairment. Alzheimers Dement 2008 Nov;4(6):395-405 [FREE Full text] [doi: 10.1016/j.jalz.2008.07.004] [Medline: 19012864]
17. Dodge HH, Mattek NC, Austin D, Hayes TL, Kaye JA. In-home walking speeds and variability trajectories associated with mild cognitive impairment. Neurology 2012 Jun 12;78(24):1946-1952 [FREE Full text] [doi: 10.1212/WNL.0b013e318259e1de] [Medline: 22689734]
18. Petersen RC, Morris JC. Mild cognitive impairment as a clinical entity and treatment target. Arch Neurol 2005 Jul;62(7):1160-3; discussion 1167. [doi: 10.1001/archneur.62.7.1160] [Medline: 16009779]
19. Toh X, Tan HX, Liang H, Tan H. Elderly Medication Adherence Monitoring With the Internet of Things. In: Proceedings of the 2016 IEEE International Conference on Pervasive Computing and Communication Workshops. 2016 Presented at: PerCom Workshops'16; March 14-18, 2016; Sydney, NSW, Australia. [doi: 10.1109/PERCOMW.2016.7457133]

20. Ng BT, Tan H, Tan HX. Managing Sensor Systems for Early Detection of Mild Cognitive Impairment in Community Elderly: Lessons Learned and Future Work. In: Proceedings of the IRC Conference on Science Engineering Technology 2017. 2017 Presented at: IRC-SET'17; August 10-11, 2017; Singapore.
21. Dlugaj M, Weinreich G, Weimar C, Stang A, Dragano N, Wessendorf TE, Heinz Nixdorf Recall Study Investigative Group. Sleep-disordered breathing, sleep quality, and mild cognitive impairment in the general population. J Alzheimers Dis 2014;41(2):479-497. [doi: 10.3233/JAD-132132] [Medline: 24643134]
22. Palmer K, Mitolo M, Burgio F, Meneghello F, Venneri A. Sleep disturbance in mild cognitive impairment and association with cognitive functioning. A case-control study. Front Aging Neurosci 2018;10:360 [FREE Full text] [doi: 10.3389/fnagi.2018.00360] [Medline: 30473661]
23. Boise L, Wild K, Mattek N, Ruhl M, Dodge H, Kaye J. Willingness of older adults to share data and privacy concerns after exposure to unobtrusive in-home monitoring. Gerontechnology 2013;11(3):428-435 [FREE Full text] [doi: 10.4017/gt.2013.11.3.001.00] [Medline: 23525351]

Abbreviations

CDR: Clinical Dementia Rating
FS: Friendship Scale
GDS: Geriatric Depression Scale
HC: healthy cognition
JAS: Jurong Ageing Study
MCI: mild cognitive impairment
MMSE: Mini-Mental State Examination
MoCA: Montreal Cognitive Assessment
PIR: passive infrared
PSQI: Pittsburgh Sleep Quality Index
SDS: Self-Rating Depression Scale

Edited by G Eysenbach; submitted 31.10.19; peer-reviewed by LL Ng, KH Tay, Y Zhou, Q Fang; comments to author 25.11.19; revised version received 14.12.19; accepted 26.01.20; published 05.05.20.

Please cite as:
Rawtaer I, Mahendran R, Kua EH, Tan HP, Tan HX, Lee TS, Ng TP
Early Detection of Mild Cognitive Impairment With In-Home Sensors to Monitor Behavior Patterns in Community-Dwelling Senior Citizens in Singapore: Cross-Sectional Feasibility Study
J Med Internet Res 2020;22(5):e16854
URL: https://www.jmir.org/2020/5/e16854
doi:10.2196/16854
PMID:32369031

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org